TESTING THE COINTEGRATING RANK WHEN THE ERRORS ARE UNCORRELATED BUT NONINDEPENDENT

Hamdi Raïssi

TESTING THE COINTEGRATING RANK WHEN THE ERRORS ARE UNCORRELATED BUT NONINDEPENDENT

HAMDI RAÏSSI,* *Université Lille 3*

Abstract

We study the asymptotic behaviour of the reduced rank estimator of the cointegrating space and adjustment space for vector error correction time series models with nonindependent innovations. It is shown that the distribution of the adjustment space can be quite different for models with iid innovations and models with nonindependent innovations. It is also shown that the likelihood ratio test remains valid when the assumption of iid Gaussian errors is relaxed. Monte Carlo experiments illustrate the finite sample performance of the likelihood ratio test using various kinds of weak error processes.

Keywords: Cointegration, reduced rank regression, likelihood ratio test, strong mixing condition, vector error correction model.

1. Introduction

Multivariate processes are often used in econometric applications because they allow to understand the interactions between different variables. In order to describe long run economic relationships, the cointegration theory has been developed by Granger (1981), Engle and Granger (1987), Ahn and Reinsel (1990). This theory postulates that, in some cases, a stationary process of lower dimension is obtained by considering linear combinations of the components of a multivariate nonstationary process. The number of independent linear combinations is the cointegrating rank and is an important piece of information for the analysis of economic data.

The dominant test for the cointegrating rank is the likelihood ratio (LR) test developed by Johansen (1988, 1991), Perron and Campbell (1993), Lütkepohl and Saikkonen (1999) in the framework of vector error correction models (VECM). For the cointegration analysis, the errors terms are generally supposed to be independent and identically distributed (iid). When applied to economic data (see for instance Johansen and Juselius (1990), Clements and Hendry (1996) or Trenkler (2003)), this iid assumption seems too restrictive because macroeconomic time series often exhibit conditional heteroscedasticity and/or other forms of nonlinearity.

Rahbek, Hansen and Dennis (2002) studied the effect of ARCH innovations on the LR test. An important output of their work is that the LR test remains valid when the error process is a martingale difference. However the assumption that the error process is a martingale difference precludes other forms of dependence. Indeed there exist many examples where the assumption of iid or martingale difference on the innovations is not satisfied (see for instance Francq, Roy and Zakoïan (2005) in the univariate ARMA case or Francq and Raïssi (2005) in the VAR case). The first aim of this paper is to study the validity of the LR test in a general context of uncorrelated errors.

* Postal address: Université Lille 3, EQUIPPE Universités de Lille BP 60149 Villeneuve D'Ascq Cedex, France. E-mail: hamdi.raissi@etu.univ-lille3.fr.

The second aim is to study the asymptotic behaviour of the usual estimators of the cointegration and adjustment spaces, in the general framework of VECM with uncorrelated, but possibly dependent errors. We will compare our findings to the usual iid case and results of Seo (2007) which shows in particular that the asymptotic distribution of the reduced rank estimator of the cointegrating space is robust to conditional heteroskedasticity. We will use the standard reduced rank procedure to estimate the cointegration space, relaxing the assumption of iid gaussian innovations.

The structure of the paper is as follows. In Section 2 we present the model and we derive the estimators of the parameters. In Section 3 we give the asymptotic behaviour of the LR test. In section 4 we state the consistency of the cointegration space and the adjustment space. In Section 5 Monte Carlo experiments are performed. The proofs are relegated to the appendix.

In the sequel the following notations are used. Weak convergence is denoted by $\Rightarrow$ and we denote by $\xrightarrow{P}$ the convergence in probability. For a full column rank matrix A of dimension $d \times r$ with $d > r$, we define the orthogonal complement $A_\perp$, which is a full column rank matrix of dimension $d \times (d-r)$ and such that $A'A_\perp = 0$. The symbol $\otimes$ denotes the usual Kronecker product and $\mathrm{vec}(A)$ denotes the vector obtained by stacking the column of the matrix A. We denote by $tr(B)$ the trace of a square matrix B. We denote by $[m]$ the integer part of a given real m.

2. Characterization of the model

We consider the following VECM with linear trend

$$\Delta X_t = \Pi_0 X_{t-1} + \sum_{i=1}^{p-1} \Gamma_{0i} \Delta X_{t-i} + \mu_{o0} + \mu_{o1} t + \epsilon_t \tag{2.1}$$

where μ_{o0} and μ_{o1} are d-dimensional parameter vectors. The process (ϵ_t) is usually assumed iid with mean zero and positive definite covariance matrix Σ_ϵ. In the sequel we will consider a weaker assumption for the error process. The Γ_{0i}, $i \in \{1, ..., p-1\}$, are $d \times d$ short run parameters matrices. By convention the sum vanishes in (2.1) when $p = 1$. The following assumption gives us the general framework of our study.

Assumption A1 (Cointegration and restriction on the trend parameters)

(a) The matrix Π_0 is of rank r_0 $(0 \leq r_0 < d)$. If $r_0 > 0$ then Π_0 can be written as $\Pi_0 = \alpha_0 \beta_0'$ where α_0 and β_0 are full column rank matrices of dimension $d \times r_0$.

(b) The autoregressive polynomial $A(z) = (1-z)I_d - \Pi_0 z - \sum_{i=1}^{p-1} \Gamma_{0i}(1-z)z^i$, is such that $\mid A(z) \mid = 0$ implies that $\mid z \mid > 1$ or $z = 1$.

(c) The matrix $\alpha_{0\perp}' \Gamma_0 \beta_{0\perp}$ is of full rank $d - r_0$, where $\Gamma_0 = I_d - \sum_{i=1}^{p-1} \Gamma_{0i}$.

(d) The vector μ_{o1} is such that $\mu_{o1} = -\alpha_0 \tau_0$, where $\tau_0 \neq 0$ is an r_0-dimensional vector.

Note that if $r_0 = 0$ the relation (2.1) is a vector autoregressive model for the process (ΔX_t). Condition (d) is the less restrictive condition on the parameters of the deterministic part of (2.1) which allows for trending behaviour for (X_t). Indeed under **A1**, from Granger's representation theorem, the solution of (2.1) has the following representation

$$X_t = C\sum_{i=1}^{t}\epsilon_i + \rho_{o1}t + \rho_{o0} + Y_t + A, \tag{2.2}$$

where $C = \beta_{0\perp}(\alpha_{0\perp}'\Gamma_0\beta_{0\perp})^{-1}\alpha_{0\perp}'$. The term A depends on initial values and is such that $\beta_0'A = 0$. The stationary process (Y_t) is of the form

$$Y_t = \sum_{i=0}^{\infty}\varphi_{0i}\epsilon_{t-i},$$

where $C(z) = \sum_{i=0}^{\infty}\varphi_{0i}z^i$ is convergent for $\mid z \mid\leq 1+\delta$, for some $\delta > 0$. Note that (2.2) implies that (X_t) is an $I(1)$ process. From (a) and (d) we can write (2.1) as

$$\Delta X_t = \nu_0 + \alpha_0\beta_0^{\prime *}Z_{1t} + \sum_{i=1}^{p-1}\Gamma_{0i}\Delta X_{t-i} + \epsilon_t \tag{2.3}$$

where $Z_{1t} = (X_{t-1}', -t+1)'$ and $\beta_0^* = (\beta_0', \tau_0)'$. The d-dimensional vector of constants ν_0 and the r_0-dimensional vector τ_0 are functions of the parameters in (2.1). Note that in (2.2) the vector ρ_{o1} is such that $\beta_0'\rho_{o1} = \tau_0$. Then it can be seen from (2.2) that $(\beta_0'X_t - E(\beta_0'X_t))$ is trend stationary and the r_0-dimensional process $(\beta_0^{\prime *}Z_{1t} - E(\beta_0^{\prime *}Z_{1t}))$ is stationary. We say in this case that the cointegrating rank is r_0. In this study we test, for some r $(0 \leq r < d)$, the null hypothesis

$$H_0 : r_0 = r \qquad \text{vs.} \qquad H_1 : r_0 > r.$$

Note that in (2.3) the parameters α_0, β_0 and τ_0 are not identified. Indeed for a given α_{01}, β_{01}, and since we assumed that these matrices have full rank, we can take any non singular matrix ζ of dimension $r_0\times r_0$ such that $\beta_{02} = \beta_{01}\zeta$ and $\alpha_{02} = \alpha_{01}(\zeta')^{-1}$ will give the same matrix Π_0. To get rid of this problem one can consider the following normalization

$$\beta_{0c}^* = (\beta_{0c}', \tau_{0c})' = ((\beta_0(c'\beta_0)^{-1})', (\beta_0'c)^{-1}\tau_0)' \quad \text{and} \quad \alpha_{0c} = \alpha_0\beta_0'c,$$

where the dimensional $d\times r_0$ matrix c is such that $c'\beta_0$ has full rank. This normalization ensures identifiability in the sense that we have $\beta_{01c} = \beta_{02c}$. To see this, note that

$$\begin{aligned} c'\beta_{01c} = c'\beta_{02c} = I_{r_0} \quad \Rightarrow \quad & c'\beta_{01}(c'\beta_{01})^{-1} = c'\beta_{01}\zeta(c'\beta_{01}\zeta)^{-1} \\ & \Rightarrow c'\beta_{01}\left[(c'\beta_{01})^{-1} - \zeta(c'\beta_{01}\zeta)^{-1}\right] = 0. \end{aligned} \tag{2.4}$$

Then since $c'\beta_{01}$ is a full rank matrix, this implies that

$$(c'\beta_{01})^{-1} - \zeta(c'\beta_{01}\zeta)^{-1} = 0. \tag{2.5}$$

Multiplying (2.5) by β_{01} on the left, we obtain $\beta_{01c} = \beta_{02c}$. Once the parameter β_{0c} is identified, it is easy to see that α_{0c} and τ_{0c} are also identified. It should be also noted that the cointegration space and the adjustment space, that is the spaces spanned by respectively β_{0c} and α_{0c}, do not depend on the choice of the matrix c.

In general the assumption that (ϵ_t) is iid gaussian may appear to be too strong. Indeed it is questionable to assume that a linear combination of $X_{t-1}, \dots, X_{t-p}$ is the best predictor of X_t. In addition note that, from a practical point of view, the order p is often identified using tests that are only based on the autocorrelations of (ϵ_t). For instance let us consider the daily exchange rates of U.S. Dollars to one British Pound and of U.S. Dollars to one Euro from January 2, 2001 to April 12, 2007. The length of the series is $T = 1578$. The analyzed data are plotted in Figure 7.10. We adjusted the model (2.1) to the series with $r_0 = 1$ and $p = 2$ using the software JMulTi. Figures 7.11-7.12 display the autocorrelations and crosscorrelations of the residuals. Figures 7.13-7.14 display the autocorrelations and crosscorrelations of the squared component of the residuals. In view of Figures 7.11-7.12 the hypothesis of uncorrelated errors seems plausible. Indeed most of the autocorrelations and crosscorrelations are inside the 5% significance limits. However since many autocorrelations and crosscorrelations are outside the 5% significance limits in Figures 7.13-7.14, the hypothesis of independent errors is clearly rejected.

Rahbek *et al* (2002) considered VECM with martingale difference innovations. In our framework we will consider a more general assumption allowing for a large class of error processes.

Assumption A2 The error process (ϵ_t) is strictly stationary and such that $Cov(\epsilon_t, \epsilon_{t-h}) = 0$ for all t $\in \mathbb{Z}$ and all $h \neq 0$.

Such error processes are commonly named weak white noise. Note that Granger's representation theorem still holds when the assumption of iid gaussian innovations is replaced by **A2**. The following are examples of error processes which verify **A2** but are not iid.

Example 2.1. Consider the process (ϵ_t) defined by the relation

$$\epsilon_t = a_t + \Phi\{\epsilon_{t-1} \odot a_t\}, \tag{2.6}$$

where $\odot$ denotes the Hadamard product, (a_t) is a d-dimensional iid centered process such that $\mid E(a_{it}a_{jt}) \mid \leq 1$, and the matrix Φ is diagonal of dimension $d \times d$ and such that $\mid \Phi_{ii} \mid < 1$. Taking $\Phi^0 = I_d$, the equation (2.6) has a stationary solution of the form $\epsilon_t = \sum_{i=0}^{\infty} \Phi^i a_{t-i} \odot \cdots \odot a_t$. It is easy to see that the ϵ_t's are uncorrelated. However

$$Cov(\epsilon_{it}^2, \epsilon_{it-1}^2) = E(a_{it}^2) Cov((1 + \Phi_{ii}\epsilon_{it-1})^2, \epsilon_{it-1}^2) \neq 0,$$

in general, showing that the process (ϵ_t) is not iid.

Example 2.2. The univariate all-pass models (see for instance Breidt, Davis and Trindade (2001)) constitute an important class which can be extended to the multivariate case. Assume that the process (ϵ_t) is the unique solution to the following equation

$$\epsilon_t - \phi_{01}\epsilon_{t-1} - \cdots - \phi_{0q}\epsilon_{t-q} = w_t + \phi_{0q-1}\phi_{0q}^{-1} w_{t-1} + \cdots + \phi_{01}\phi_{0q}^{-1} w_{t-q+1} - \phi_{0q}^{-1} w_{t-q},$$

where $\phi(z) = I_d - \phi_{01} z \cdots - \phi_{0q} z^q$ is such that $\phi(z) \neq 0$ for $\mid z \mid \leq 1$. The centered process (w_t) is iid with variance Σ_w. Assume also that the matrices $\phi_{01}, \dots, \phi_{0q}$ are diagonal. Writing the spectral density for each component (ϵ_{it}), it can be shown that the process (ϵ_t) is uncorrelated (see Andrews, Davis and Breidt (2006)). However if y_0 is not gaussian the process (ϵ_t) is not independent. To see this consider the following bivariate simple example

$$\epsilon_t - \phi\epsilon_{t-1} = w_t - \phi^{-1}w_{t-1}$$

where $\phi = \begin{pmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{pmatrix}$ and $\mid \phi_1 \mid < 1$, $\mid \phi_2 \mid < 1$. Let us introduce $\vartheta_t = \epsilon_{1t} - \phi_1\epsilon_{1t-1}$. Since (ϵ_t) is uncorrelated, the process (ϑ_t) follows an causal $MA(1)$. Then we have $\epsilon_{1t} = \sum_{i\geq 0}\phi^i\vartheta_{t-i}$. Straightforward computations show that $E(\epsilon_{1t}\vartheta_{t-1}^2) = E[\epsilon_{1t}(\epsilon_{1t-1}-\epsilon_{1t-2})^2] = Ew_t^3(1-\phi_1^{-2})(1+\phi_1)$ and $E(\epsilon_{1t}\vartheta_{t-1}^3) = E[\epsilon_{1t}(\epsilon_{1t-1}-\epsilon_{1t-2})^3] = (Ew_t^4-3)(1-\phi_1^{-2})^2\phi_1$. Using the fact that ϑ_{t-1} belongs to the σ-field generated by $\{\epsilon_{1u}, \quad u<t\}$, we have $E\{\vartheta_{t-1}^2 E(\epsilon_{1t}\mid\epsilon_{1t-1},\cdots)\}\neq 0$ for $Ew_t^3\neq 0$ and $E\{\vartheta_{t-1}^3 E(\epsilon_{1t}\mid \epsilon_{1t-1},\cdots)\}\neq 0$ for $Ew_t^4\neq 0$. Thus the (ϵ_t) process is not a martingale difference in general.

2.1. Derivation of the quasi maximum likelihood (QML) estimators

Now we turn to the derivation of the QML estimators of α_{0c} and β_{0c}^*. We use here the QML method because we assume that the errors terms are uncorrelated but not necessary gaussian independent. Note that the estimation procedure we will describe is performed under H_0. In the framework of the VECM we shall see that the methodology in Johansen (1988,1991) in the iid case remains valid under uncorrelated errors assumption. We will use the following notation. Let $Z_{0t} = \Delta X_t$, $Z_{2t} = (\Delta X'_{t-1},\ldots,\Delta X'_{t-p+1},1)'$, $\Psi_0 = (\Gamma_{01},\ldots,\Gamma_{0p-1},\nu_0)$ where $X_t = 0$ for $t\leq 0$. The expression (2.3) becomes with these notations

$$Z_{0t} = \alpha_{0c}\beta_{0c}^{\prime *}Z_{1t} + \Psi_0 Z_{2t} + \epsilon_t. \tag{2.7}$$

Here we can remark that since X_t is $I(1)$ then the processes Z_{0t} and Z_{2t} are stationary. Using (2.7) and given the observations $X_1,\ldots,X_T$ we write the quasi log-likelihood as follows

$$\log L(\Psi,\alpha_c,\beta_c,\Sigma_\epsilon) = -\frac{1}{2}T\log\mid\Sigma_\epsilon\mid$$
$$-\frac{1}{2}tr\left\{\sum_{t=1}^T \Sigma_\epsilon^{-1}(Z_{0t}-\alpha_c\beta_c^{\prime *}Z_{1t}-\Psi Z_{2t})(Z_{0t}-\alpha_c\beta_c^{\prime *}Z_{1t}-\Psi Z_{2t})'\right\},$$

where

$$\beta_c^* = (\beta_c',\tau_c)' = ((\beta(c'\beta)^{-1})',(\beta'c)^{-1}\tau)' \quad \text{and} \quad \alpha_c = \alpha\beta'c.$$

The maximum likelihood estimation method for the VECM with uncorrelated errors implicates several steps. We first estimate the parameters in the matrix Ψ_0 and obtain

$$\hat{\Psi}(\alpha_c,\beta_c^*) = M_{02}M_{22}^{-1} - \alpha_c\beta_c^{\prime *}M_{12}M_{22}^{-1}$$

where

$$M_{ij} = T^{-1}\sum_{t=1}^T Z_{it}Z'_{jt}.$$

Now defining by R_{0t} and R_{1t} the residuals of respectively the regressions of Z_{0t} and Z_{1t} on Z_{2t}, we get the concentrated log-likelihood

$$\begin{aligned}\log L(\alpha_c, \beta_c^*, \Sigma_\epsilon) &= -\frac{1}{2} T \log \mid \Sigma_\epsilon \mid \\ &\quad -\frac{1}{2} tr \left\{ \sum_{t=1}^{T} \Sigma_\epsilon^{-1} (R_{0t} - \alpha_c \beta_c^{\prime *} R_{1t})(R_{0t} - \alpha_c \beta_c^{\prime *} R_{1t})' \right\} \end{aligned} \tag{2.8}$$

where

$$R_{0t} = Z_{0t} - M_{02} M_{22}^{-1} Z_{2t} \quad \text{and} \quad R_{1t} = Z_{1t} - M_{12} M_{22}^{-1} Z_{2t}.$$

Since the R_{1t}'s are the residuals of the regression of the Z_{1t}'s on the Z_{2t}'s, and noting that the process (Z_{1t}) is $I(1)$ and the process (Z_{2t}) is $I(0)$, then the process (R_{1t}) is $I(1)$. The expression of the concentrated log-likelihood corresponds to the regression equation

$$R_{0t} = \alpha_{0c} \beta_{0c}^{\prime *} R_{1t} + \tilde{\epsilon}_t, \tag{2.9}$$

so that we obtain the following unfeasible estimators of α_{0c} and Σ_ϵ in (2.9) by ordinary least squares

$$\begin{aligned}\hat{\alpha}_c(\beta_{0c}^*) &= S_{01} \beta_{0c}^* (\beta_{0c}^{\prime *} S_{11} \beta_{0c}^*)^{-1}, \\ \hat{\Sigma}_\epsilon(\beta_{0c}^*) &= S_{00} - \hat{\alpha}_c(\beta_{0c}^*)(\beta_{0c}^{\prime *} S_{11} \beta_{0c}^*) \hat{\alpha}_c'(\beta_{0c}^*)\end{aligned} \tag{2.10}$$

where

$$S_{ij} = T^{-1} \sum_{t=1}^{T} R_{it} R_{jt}'.$$

Note that replacing α_c and Σ_ϵ by their estimates in (2.8) we write

$$\log L(\hat{\alpha}(\beta_c^*), \beta_c^*, \hat{\Sigma}_\epsilon(\beta_c^*)) = -\frac{1}{2} T \log \mid \hat{\Sigma}_\epsilon(\beta_c^*) \mid -\frac{1}{2} dT.$$

Finally the parameters in β_{0c}^* can be estimated using the results of the well known reduced rank method of Anderson (1951). In this end we shall minimize the following expression

$$\mid \hat{\Sigma}_\epsilon(\beta_c^*) \mid = \mid S_{00} - S_{01} \beta_c^* (\beta_c^{\prime *} S_{11} \beta_c^*)^{-1} \beta_c^{\prime *} S_{10} \mid .$$

Using the relation

$$\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = \mid A_{11} \mid \mid A_{22} - A_{21} A_{11}^{-1} A_{12} \mid = \mid A_{22} \mid \mid A_{11} - A_{12} A_{22}^{-1} A_{21} \mid,$$

we find

$$\mid S_{00} - S_{01} \beta_c^* (\beta_c^{\prime *} S_{11} \beta_c^*)^{-1} \beta_c^{\prime *} S_{10} \mid = \mid S_{00} \mid \frac{\mid \beta_c^{\prime *} (S_{11} - S_{10} S_{00}^{-1} S_{01}) \beta_c^* \mid}{\mid \beta_c^{\prime *} S_{11} \beta_c^* \mid}.$$

Under the null hypothesis and using Lemma 7.1 the expression $\mid \beta_c^{\prime *} (S_{11} - S_{10} S_{00}^{-1} S_{01}) \beta_c^* \mid / \mid \beta_c^{\prime *} S_{11} \beta_c^* \mid$ is minimized for the following normalized expression

$$\hat{\beta}_c^* = (\hat{\beta}_c', \hat{\tau}_c)' = ((\hat{\beta}(c'\hat{\beta})^{-1})', ((\hat{\beta}'c)^{-1} \hat{\tau}))',$$

where

$$\hat{\beta}^* = (\hat{\beta}', \hat{\tau})' = S_{11}^{-\frac{1}{2}}(v_1, \ldots, v_r)$$

and $v_1, \ldots, v_r$ are eigenvectors corresponding to the r largest solutions $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_r$ of the eigenvalue problem

$$\mid \lambda I - S_{11}^{-\frac{1}{2}} S_{10} S_{00}^{-1} S_{01} S_{11}^{-\frac{1}{2}} \mid = 0. \tag{2.11}$$

In addition the matrix $c'\hat{\beta}$ is of full rank. We obtain $\hat{\alpha}_c = S_{01}\hat{\beta}_{0c}^*(\hat{\beta}_{0c}'^* S_{11}\hat{\beta}_{0c}^*)^{-1}$. Noting that we have $\mid \hat{\Sigma}_\epsilon(\hat{\beta}_c^*) \mid = \prod_{i=1}^r (1 - \hat{\lambda}_i)$, the likelihood ratio test for r is given by

$$Q_r^{-\frac{2}{T}} = \frac{\prod_{i=1}^r (1 - \hat{\lambda}_i)}{\prod_{i=1}^d (1 - \hat{\lambda}_i)} = \prod_{i=r+1}^d (1 - \hat{\lambda}_i)^{-1}.$$

Then to test the null hypothesis, we consider the LR test statistic

$$-2 \log Q_r = -T \sum_{i=r+1}^d \log(1 - \hat{\lambda}_i),$$

where $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_d$ are the d greater solutions of the eigenvalue problem (2.11). In the next section we will study the asymptotic behaviour of the LR test statistic.

3. Asymptotic properties of the LR statistic

To state the main results of the paper, the assumption that the process (ϵ_t) is uncorrelated is not enough. Indeed we have to control the serial dependence of the process (ϵ_t). To this end we introduce the mixing coefficients $\alpha_\xi(h)$ for a given stationary process (ξ_t)

$$\alpha_\xi(h) = \sup_{A \in \sigma(\xi_u, u \leq t), B \in \sigma(\xi_u, u \geq t+h)} \left| P(A \cap B) - P(A)P(B) \right|,$$

which measures the temporal dependence of the process (ξ_t). Define $\|\xi_t\|_q = (E\|\xi_t\|^q)^{1/q}$, where $\|.\|$ denotes the Euclidean norm. Then we need to make the following assumption on the process (ϵ_t).

Assumption A3 The process (ϵ_t) satisfies $\|\epsilon_t\|_{2+\nu+\eta} < \infty$ and the mixing coefficients of the process (ϵ_t) are such that $\sum_{h=0}^\infty \{\alpha_\epsilon(h)\}^{\nu/(2+\nu)} < \infty$ for some $\nu > 0$ and $\eta > 0$.

Note that the kind of dependence induced by **A3** is mild for the error process (ϵ_t). The following proposition gives us the asymptotic distribution of the LR test statistic.

Proposition 3.1. *Under* **A1**, **A2** *and* **A3**, *the LR test statistic has the same asymptotic distribution as in the iid gaussian case, that is*

$$-2 \log Q_{r_0} \Rightarrow tr \left\{ \left[\int_0^1 F(dB)' \right]' \left[\int_0^1 FF'du \right]^{-1} \left[\int_0^1 F(dB)' \right] \right\}, \tag{3.1}$$

where B is a standard $d-r_0$ dimensional Brownian motion, and the components F_i of F are given by

$$\begin{aligned} F_i(u) &= B_i(u) - \bar{B}_i \quad i = 1, \ldots, d-r_0, \\ F_{d-r_0+1}(u) &= u - \frac{1}{2}, \end{aligned}$$

and $\bar{B}_i = \int_0^1 B_i(u)du$.

The same result was found by Rahbek et *al* (2002) under the assumption that the error process (ϵ_t) is a martingale difference and in the framework of VECM without deterministic terms. A consequence of Proposition 3.1 is that the results for testing the cointegrating rank using the LR test statistic can be directly extended from the usual iid gaussian assumption on the error process. Then we can use the same critical values as in the iid case to test the cointegrating rank (see Johansen (1995), Table 15.4). We reject the null hypothesis if $-2\log Q_r > \varsigma$ for a given quantile ς of the distribution given in (3.1). Therefore, following the Johansen procedure for selecting the cointegrating rank, we apply successively this test to $r = 0, 1, 2, \ldots, d-1$ until we obtain $-2\log Q_r < \varsigma$. Note that if $\tau_0 = 0$, we use a different test statistic and a different limit distribution is obtained in this case. In the next section we will study the asymptotic behaviour of the QML estimators.

4. Asymptotic properties of the QML estimators

In this section we suppose that the cointegrating rank is well identified and only consider estimates of β_{0c}^* with dimension $(d \times r_0)$. In the sequel we will denote by $W(u)$ the d-dimensional brownian motion of variance Σ_ϵ and define $\bar{W} = \int_0^1 W(u)du$. We also define the matrix $\bar{\beta}_0 = \beta_0(\beta_0'\beta_0)^{-1}$. The following Proposition gives the asymptotic behaviour of β_{0c}^*.

Proposition 4.1. *Under **A1**, **A2** and **A3**, $T(\hat{\beta}_c - \beta_{0c})$ has the same asymptotic distribution as in the iid gaussian case that is*

$$T(\hat{\beta}_c - \beta_{0c}) \Rightarrow (I_d - \beta_{0c}c')\bar{\beta}_{0\perp}\left[\int_0^1 G_{1.2}G_{1.2}'du\right]^{-1}\int_0^1 G_{1.2}(dV_\alpha)' \qquad (4.1)$$

where

$$G(u) = \begin{pmatrix} \bar{\beta}_{0\perp}'C(W(u) - \bar{W}) \\ -u + \frac{1}{2} \end{pmatrix} = \begin{pmatrix} G_1(u) \\ G_2(u) \end{pmatrix},$$

$$G_{1.2} = G_1 - \left(\int_0^1 G_1G_2du\right)\left(\int_0^1 G_2G_2du\right)^{-1} G_2,$$

and

$$V_\alpha = (\alpha_{0c}'\Sigma_\epsilon^{-1}\alpha_{0c})^{-1}\alpha_{0c}'\Sigma_\epsilon^{-1}W$$

is independent of G. Then $\hat{\beta}_c$ is asymptotically distributed as mixture normal with variance

$$(I_d - \beta_{0c}c')\bar{\beta}_{0\perp}\left[\int_0^1 G_{1.2}G_{1.2}'du\right]^{-1}(I_d - c\beta_{0c}') \otimes (\alpha_{0c}'\Sigma_\epsilon^{-1}\alpha_{0c})^{-1}. \qquad (4.2)$$

Moreover we have

$$\hat{\tau}_c = \tau_{0c} + O_p(T^{-\frac{3}{2}}). \tag{4.3}$$

Seo (2007) also found that the asymptotic distribution of the reduced rank estimator is not changed when the errors are conditionally heteroscedastic.

It is interesting to note that the results of Proposition 3.1 and Proposition 4.1 remain valid considering $\Upsilon_t = ((\beta_0'^* Z_{1t})', Z_{0t}')'$ and replacing **A3** by the following assumption.

Assumption A3' The process (Υ_t) satisfies $\|\Upsilon_t\|_{2+\nu+\eta} < \infty$, moreover the mixing coefficients of the process (Υ_t) are such that

$$\sum_{h=0}^{\infty} \{\alpha_\Upsilon(h)\}^{\nu/(2+\nu)} < \infty \quad \text{for} \quad \text{some} \quad \nu > 0 \quad \text{and} \quad \eta > 0. \tag{4.4}$$

However assumptions **A3** and **A3'** are not equivalent. Note that using **A3'** we consider $I(0)$ transformations of the process (X_t), that is $\beta_0'^* Z_{1t}$ and Z_{0t}, so that we are able to use the theory of stationary mixing processes in our framework. Note also that the summability condition (4.4), implies that $((\beta_0'^* Z_{1t})', Z_{0t}')'$ and $((\beta_0'^* Z_{1t+h})', Z_{0t+h}')'$ are asymptotically independent while it is assumed there exist long-run relations between the components of X_t. A simple illustration of the kind of processes we consider is given by the following bivariate $I(1)$ process $X_t = (X_{1\,t}, X_{2\,t})$ such that

$$\begin{aligned} X_{1\,t} &= \nu_1 \textstyle\sum_{i=1}^t \epsilon_{0i} + \nu_1 t + \epsilon_{1t} \\ X_{2\,t} &= \nu_2 \textstyle\sum_{i=1}^t \epsilon_{0i} + \nu_2 t + \epsilon_{2t} \end{aligned}$$

where the process $(\epsilon_{0t}, \epsilon_{1t}, \epsilon_{2t})$ is a mixing process, and $\nu_1 \neq 0$ and $\nu_2 \neq 0$. Here taking $\beta_0 = (\nu_2, -\nu_1)$ it is clear that the process $(\beta_0' X_t, \Delta X_t)$ is mixing.

In order to state the consistency of the estimator of α_{0c}, we have to introduce the following notations. Let us define

$$\beta_{0c}'^* \tilde{R}_{1t} = \beta_{0c}'^* Z_{1t} - \beta_{0c}'^* \bar{M}_{12} \bar{M}_{22}^{-1} Z_{2t} \tag{4.5}$$

where

$$\beta_{0c}'^* \bar{M}_{12} = \lim_{T\to\infty} \beta_{0c}'^* M_{12} \quad \text{and} \quad \bar{M}_{22} = \lim_{T\to\infty} M_{22}.$$

The existence of these limits is ensured by the ergodic theorem since the processes $(\beta_{0c}'^* Z_{1t})$ and (Z_{2t}) are stationary ergodic. Define the matrix $\Sigma_c = E(\beta_{0c}'^* R_{1t} R_{1t}' \beta_{0c}^*) = Var(\beta_{0c}'^* Z_{1t}) - E(\beta_{0c}'^*(Z_{1t} - \bar{Z}_1)(\tilde{Z}_{2t} - \bar{\tilde{Z}}_2)')Var(\tilde{Z}_{2t})^{-1} E((\tilde{Z}_{2t} - \bar{\tilde{Z}}_{2t})(Z_{1t} - \bar{Z}_1)' \beta_{0c}^*)$ where $\tilde{Z}_{2t} = (Z_{0t-1}', \dots, Z_{0t-p+1}')'$. We also need to consider the following assumption which strengthens **A3**.

Assumption A4 The process (ϵ_t) satisfies $\|\epsilon_t\|_{4+2\nu} < \infty$ and the mixing coefficients of the process (ϵ_t) are such that $\sum_{h=0}^{\infty} \{\alpha_\epsilon(h)\}^{\nu/(2+\nu)} < \infty \quad \text{for} \quad \text{some} \quad \nu > 0.$

The following Proposition give us the asymptotic behaviour of the estimator of α_{0c}.

Proposition 4.2. *Under* **A1**, **A2** *and* **A4**, *the expression* $T^{\frac{1}{2}}vec(\hat{\alpha}_c - \alpha_{0c})$ *has the following asymptotic distribution which is different from that of the usual iid gaussian case,*

$$T^{\frac{1}{2}}vec(\hat{\alpha}_c - \alpha_{0c}) \Rightarrow \mathcal{N}(0, \Sigma_\alpha) \tag{4.6}$$

where

$$\Sigma_\alpha = \sum_{h=-\infty}^{\infty} E\left\{\Sigma_c^{-1}\beta_{0c}^{\prime *}\tilde{R}_{1t}\tilde{R}_{1t-h}^{\prime}\beta_{0c}^{*}\Sigma_c^{-1} \otimes \epsilon_t\epsilon_{t-h}^{\prime}\right\}.$$

In the iid gaussian case the asymptotic variance is given by

$$\Sigma_\alpha = \Sigma_c^{-1} \otimes \Sigma_\epsilon,$$

so that in this case (4.6) corresponds to the result in Johansen (1995, Theorem 13.3 p 183). We also can obtain the result of Proposition 4.2 replacing **A4** by the following assumption.

Assumption A4' The process (Υ_t) satisfies $\|\Upsilon_t\|_{4+2\nu} < \infty$, moreover the mixing coefficients of the process (Υ_t) are such that

$$\sum_{h=0}^{\infty}\{\alpha_\Upsilon(h)\}^{\nu/(2+\nu)} < \infty \quad \text{for} \quad \text{some} \quad \nu > 0.$$

Then despite the fact that the assumption of iid gaussian noise is relaxed in the estimation procedure, the estimates of α_{0c} and β_{0c}^* obtained in Section 2 are consistent.

5. Monte Carlo experiments

In this section we compare the small sample properties of the LR test in the cases of iid and dependent innovations for bivariate processes. Throughout this section the error process is normally distributed with mean zero and variance matrix I_2 in the iid case. We will consider several kinds of weak error processes. Consider the iid process $\eta_t = (\eta_{1t}, \eta_{2t})'$ such that $\eta_t \sim \mathcal{N}(0, I_2)$. We first consider a bivariate error process defined by

$$\epsilon_t = \begin{pmatrix} \eta_{1t}\eta_{1t-1}\cdots\eta_{1t-k} \\ \eta_{2t}\eta_{2t-1}\cdots\eta_{2t-k} \end{pmatrix}, \tag{5.1}$$

for some integer k. Note that the components of ϵ_t correspond to the univariate weak white noise built by Romano and Thombs (1996). The innovations process defined in (5.1) is obviously not independent. It can be shown that (ϵ_t) is a martingale difference. Note also that the error process is k-dependent, in the sense that ϵ_t and ϵ_{t-i} are dependent for $i \leq k$ and independent for $i > k$.

In order to illustrate the effect of ARCH innovations on the LR test statistic we consider the model with constant correlation proposed by Jeantheau (1998). In our simulations the process (ϵ_t) follows the DGP given by

$$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} = \begin{pmatrix} \sigma_{1t} & 0 \\ 0 & \sigma_{2t} \end{pmatrix}\begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} \tag{5.2}$$

where

$$\begin{pmatrix} \sigma_{1t}^2 \\ \sigma_{2t}^2 \end{pmatrix} = \begin{pmatrix} 0.1 \\ 0.1 \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} \epsilon_{1t-1}^2 \\ \epsilon_{2t-1}^2 \end{pmatrix}.$$

The elements a_{11}, a_{12}, a_{21} and a_{22} are supposed to be positive. In addition we suppose that the stationarity conditions hold (see Jeantheau (1998) for more details). In this case the process (ϵ_t) is a martingale difference and presents conditional heteroscedasticity.

The third weak error process follows an all-pass model of Example 2.2 defined by

$$\epsilon_t - \phi\epsilon_{t-1} = w_t - \phi^{-1}w_{t-1}, \quad \text{where} \quad \phi = \begin{pmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{pmatrix} \tag{5.3}$$

and ϕ_1, ϕ_2 are real and such that $\mid \phi_1 \mid < 1$, $\mid \phi_2 \mid < 1$. The terms w_t are defined by $w_t = y_{2t} \odot y_{2t-1}$, where (y_t) is iid $\mathcal{N}(0, I_2)$. Note that the process (w_t) is iid but non gaussian. Contrary to the first and second case, the innovation process is not in general a martingale difference.

5.1. Empirical size

We simulated $n = 1000$ independent trajectories of length $T = 100$ and $T = 400$ given by the following bivariate DGP

$$\begin{pmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{pmatrix} = \begin{pmatrix} \pi_1 & e\pi_1 \\ \pi_2 & e\pi_2 \end{pmatrix} \begin{pmatrix} X_{1t-1} \\ X_{2t-1} \end{pmatrix} - \theta \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} (t-1) + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \tag{5.4}$$

where π_1, π_2, e and θ are real. The true cointegrating rank is $r_0 = 1$. Note that the conditions (b) and (c) of **A1** become in this case $-2 < e\pi_2 + \pi_1 < 0$. When the equation $\mid A(z) \mid = 0$ has two solutions, they will be denoted by $z_1 = 1$ and z_2. In the sequel, we consider tests of the hypothesis $H_0 : r_0 = 1$ at the asymptotic nominal level 5%, assuming the order $p = 1$ is known.

In Tables 1 and 2, we consider three different cases of the model (5.4) to study the behaviour of the LR test in different points of the parameter space. For the three cases we take $\pi_2 = 0.9$, $e = -1$ and $\theta = -1.5$ so that only π_1 changes. We take $\pi_1 = -0.1$ for Case 1, $\pi_1 = 0.8$ for Case 2 and $\pi_1 = -0.8$ for Case 3. For Case 1 we have $e\pi_2 + \pi_1 = -1$ and the equation $\mid A(z) \mid = 0$ has a unique solution which is equal to one. Note that when $e\pi_2 + \pi_1 = 0$, we have $z_2 = 1$ so that the process (X_t) is integrated of order higher than one. Actually $e\pi_2 + \pi_1 = 0$ corresponds to $\mid \alpha_{0\perp}' \Gamma_0 \beta_{0\perp} \mid = 0$ in condition (c) of **A1**. Case 2 is close to this limiting since we have $e\pi_2 + \pi_1 = -0.1 \approx 0$ and $z_2 \approx 1$. When $e\pi_2 + \pi_1 = -2$, we have $z_2 = -1$ so that the condition (b) of **A1** is not satisfied. Case 3 is close to this limiting situation since we have $e\pi_2 + \pi_1 = -1.7 \approx -2$ and $z_2 \approx -1$. We will consider for each of these cases the white noises presented above. Recall that in the iid case the error process is normally distributed with mean zero and variance matrix I_2. For the weak white noise (5.1) we take $k = 1$. For the weak white noise (5.2) we take $a_{11} = a_{21} = 0.2$, $a_{12} = 0.1$, $a_{22} = 0.4$ and for the weak white noise (5.3) we take $\phi_1 = \phi_2 = 0.7$. In the following tables WWN stands for weak white noise, MD for martingale difference and SWN for strong white noise. The relative rejection frequencies are displayed in bold type when they are outside the 5% significant limits 3.65% and 6.35% in Tables 1 and 2.

In order to illustrate the behaviour of the LR test when the effect of the weak white noises increases, we first apply the LR test when the error process follows (5.1) with different values of k in Figure 7.1. We also apply the LR test when the error process follows the ARCH model (5.2) with $a_{21} = a_{12} = 0$ and different values of $a_{11} = a_{22}$.

The results are presented in Figure 7.2. Since we assumed that $\eta_t \sim \mathcal{N}(0, I_2)$, the moments of order two exist for $a_{11} < 1$. The existence of this moment is indicated by vertical lines. Note also that the error process is strictly stationary for $a_{11} < 3.56$. The same experiment is made for the weak white noise (5.3) with different values of $\phi_1 = \phi_2$ in Figure 7.3. These experiments are performed for Case 1. We will also study the behaviour of the the LR test for different values of the trend parameter θ for each of the noises considered above. We will take the same parameters for weak white noises (5.2) and (5.3) as in Tables 1 and 2. We will also take $k = 1$ for weak white (5.1) for these experiments. The results are presented in Figures 7.4-7.7 for $\pi_1 = -0.1$, $\pi_2 = 0.9$ and $e = -1$.

We will first interpret the results for Case 1 in the different experiments we performed. In Table 1 it emerges that the LR test is more liberal when the innovation process is a martingale difference than in the case of strong innovation for the sample $T = 100$. In addition note that from Figure 7.1 the LR test is over-rejecting for increasing values of k in the weak white noise (5.1). From Figure 7.2 the same conclusion can be made when the ARCH effect increases and the moment of order two exist. From Table 1 it seems that the LR test is more conservative by comparison to the strong case when the error process follows an all-pass model. This is confirmed from Figure 7.3 when the all-pass effect increases. In general according to the results of our experiments the LR test has some difficulties to assess the cointegrating rank for small samples when the errors are not iid.

Note however that the rejection frequencies for Case 1 in Table 2 are inside the significant limits 3.65% and 6.35%. In addition Figure 7.2 shows that the results are better for samples of size $T = 400$ than for $T = 100$ when $a_{11} < 1$. This confirms that the LR test remains valid for uncorrelated errors when $\|\epsilon_t\|_{2+\nu+\eta} < \infty$. This also confirms the result of Rahbek *et al* (2002) who showed that the LR test remains valid in the framework of martingale differences, assuming the existence of moments of order two. However the rejection frequencies increases for $a_{11} < 1$. When the moments of order two do not exist ($a_{11} > 1$), it seems that the LR test is no longer valid. Similarly Figure 7.3 clearly shows that the results are better for samples of length $T = 400$ than for samples of length $T = 100$. The same can be stated from Figure 7.1 when the dependence of the error process is not strongly marked. Note that the results for samples $T = 400$ are not better from those of samples $T = 100$ for great values of k. Then the theoretical results are beared out by the results of our experiments.

Finally from Figures 7.4-7.7 it seems that the LR test becomes more conservative for small values of the trend parameters. This could be explained by the fact that when $\theta \approx 0$ the model (5.4) resembles to a model without trend. In the case of VECM without trend one should use other critical values.

In order to interpret the results of Cases 2 and 3 recall that the parameters are close to the boundary of the parameter space in these two cases. In Case 2 the root z_2 is near the point $z = 1$, and in Case 3 the root z_2 is near the unit circle but far from the point $z = 1$. From Tables 1 and 2, it seems that the finite sample performance of the LR test is not affected too much for Case 2. Note that from Figure 7.8 the LR test is clearly more liberal in Case 2 than in Case 1 when the error process follows an all-pass model. However for Case 3, according to Tables 1 and 2 the LR test has bad performances unless when the error process follows an ARCH model. Then, for a given kind of weak white noise, the small sample properties of the LR test can change when the parameters are close to the boundary.

Now we will study the validity of the asymptotic distribution of the LR test in (3.1) when the error terms are correlated. We consider a DGP of the form (5.4) with the following correlated error process

$$\epsilon_t = \cos(0.5 \arcsin(2\delta))\eta_t + \sin(0.5 \arcsin(2\delta))\eta_{t-1}.$$

It is easy to check that $Var(\epsilon_t) = I_2$ and $Corr(\epsilon_t, \epsilon_{t-1}) = \delta I_2$. We apply the LR test based on the asymptotic critical value of level 5% to a DGP of the form (5.4) for testing the hypothesis $r_0 = 1$. Clearly from Figure (7.9) the LR test turns out to be over-rejecting when δ is far from zero. In addition the results are worst for samples $T = 400$ than for $T = 100$. Then, from the results of our experiment, we can speculate that the LR test is no longer valid when the errors are correlated. This speculation seems reasonable since it can be seen from Phillips (1988) (see also Phillips and Durlauf (1986)) that the standard results we use to prove Proposition 3.1 change when the assumption of uncorrelated errors is relaxed.

Finally we consider the following non conditionally heteroscedastic errors

$$\epsilon_t = (1 + f \times t)\eta_t \tag{5.5}$$

where f is real positive, and study the small sample properties of the LR test in this case. Similarly to the previous experiment, we apply the LR test based on the asymptotic critical value of level 5% to the bivariate DGP (5.4) testing the hypothesis $r_0 = 1$. From Table 3 the LR test seems to be too conservative in presence of heteroscedastic errors. In addition the results for samples $T = 400$ are worst than for $T = 100$, so that we can also speculate in this case that the LR test is no longer valid.

5.2. Empirical power

Now we repeat the same experiments, considering the following bivariate $AR(1)$ model written in error correction form

$$\begin{pmatrix} \Delta X_{1t} \\ \Delta X_{2t} \end{pmatrix} = \begin{pmatrix} \pi_1 & e\pi_1 \\ \pi_2 & e\pi_2 + \varpi \end{pmatrix} \begin{pmatrix} X_{1t-1} \\ X_{2t-1} \end{pmatrix} - \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} (t-1) + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \tag{5.6}$$

where we choose $\varpi \neq 0$ such that the matrix $\Pi = \begin{pmatrix} \pi_1 & e\pi_1 \\ \pi_2 & e\pi_2 + \varpi \end{pmatrix}$ is of full rank ($rk(\Pi) = 2$) and $det\,(I_d - (I_d + \Pi)z) \neq 0$ for all $\mid z \mid \leq 1$. We shall test the hypothesis $H_0 : r_0 = 1$ for each of the noises considered in Tables 1 and 2. The rejection frequencies of H_0 are displayed in Tables 4-7 for an asymptotic critical value of level 5%.

Note that for Tables 4 and 5 we simulated a model (5.6) for which we have $e\pi_2 + \pi_1 = -0.85$. From the results of Table 4 it seems that the LR test is slightly less powerful in small samples when the innovations are all-pass than when they are iid. The same can be noted in Table 6 for an error process which follows an ARCH model when the simulated model (5.6) is such that $e\pi_2 + \pi_1 = -1.8 \approx -2$. In general, from Tables 5 and 7, the power increases for samples of size $T = 400$ when the values of ϖ are not too small. Surprisingly the power decreases for small values of ϖ in Table 7.

6. Conclusion

In this work we established the consistency of the estimators of the long-run parameters β_{0c} and the adjustment parameters α_{0c} in the presence of uncorrelated but nonindependent errors. We also established the robustness of the LR test in this framework, in the sense that the LR test statistic has the same asymptotic distribution as in the iid gaussian errors case. However from the simulations results it seems that the finite sample performance of the LR test strongly depends on the kind of error process. The finite sample performance also strongly depends on the position in the parameter space. More precisely the simulations results show an important size distortion when the dependence increases or when the deterministic trend is close to zero. Similar conclusions were found by Rahbek *et al* (2002) for ARCH type errors. Note also that it appears from our experiments that the LR test is no longer valid when the errors are correlated. From these findings we can draw the conclusion that, despite the asymptotic validity of the LR test, one should use it warily when the error process is suspected to be non-independent.

7. Appendix

Lemma 7.1. *Let H and K be symmetric and positive definite matrices of dimension $d \times d$. Define the following function*

$$f(x) = \mid x'Hx \mid / \mid x'Kx \mid$$

where x is a full rank matrix of dimension $d \times r$. Define also the ordered solutions $\delta_d \geq \cdots \geq \delta_1 > 0$ of the generalized eigenvalue problem

$$\mid \delta I - K^{-\frac{1}{2}} H K^{-\frac{1}{2}} \mid = 0. \tag{7.1}$$

Then $f(x)$ is minimized among all $d \times r$ matrices by any matrix of the form

$$\hat{x} = K^{-\frac{1}{2}}(e_{i_1}, \ldots, e_{i_r}), \tag{7.2}$$

where $e_{i_1}, \ldots, e_{i_r}$ are non-collinear eigenvectors corresponding to a choice of r eigenvalues δ_{i_k} of (7.1) which are such that $\delta_{i_k} \leq \delta_r$. The minimal value is given by $\prod_{i=1}^{r} \delta_i$.

Proof of Lemma 7.1. Let a $d \times d$-dimensional matrix $l = (l_{ij})$. Using the relation

$$\log(\mid I_d + l \mid) = tr(l) + o(\parallel l \parallel^2) \quad \text{where} \quad \parallel l \parallel = \max_i \sum_{i=1}^{d} \mid l_{ij} \mid,$$

we expand the expression

$$\begin{aligned} &\log \mid (x+h)'H(x+h) \mid \\ &= \log \mid x'Hx \mid + \log \mid I + (x'Hx)^{-1}(x'Hh + h'Hx + h'Hh) \mid \\ &= \log \mid x'Hx \mid + 2tr\{(x'Hx)^{-1}(x'Hh)\} + o(\parallel h \parallel^2), \end{aligned} \tag{7.3}$$

where h is a matrix of dimension $d \times r$. Since we have

$$\log f(x) = \log \mid x'Hx \mid - \log \mid x'Kx \mid$$

and using the expression (7.3), we write the derivative of the function $\log f(x)$ at the point x in the direction h

$$\begin{aligned}\lim_{s\to 0}\frac{\log f(x+sh)-\log f(x)}{s} &= \lim_{s\to 0}\frac{2tr\{((\hat{x}'H\hat{x})^{-1}\hat{x}'H-(\hat{x}'K\hat{x})^{-1}\hat{x}'K)sh\}}{s}\\ &= 2tr\{((\hat{x}'H\hat{x})^{-1}\hat{x}'H-(\hat{x}'K\hat{x})^{-1}\hat{x}'K)h\}.\end{aligned}$$

The function $\log f(x)$ has a stationary point $\hat{x}$ if the derivative at $\hat{x}$ in the direction h is zero for all h, hence the first order condition is

$$tr\{((\hat{x}'H\hat{x})^{-1}\hat{x}'H-(\hat{x}'K\hat{x})^{-1}\hat{x}'K)h\}=0. \tag{7.4}$$

Defining $\kappa=(\hat{x}'H\hat{x})^{-1}\hat{x}'H-(\hat{x}'K\hat{x})^{-1}\hat{x}'K$ the matrix of general component κ_{ij} this condition becomes

$$\sum_{i=1}^{r}\sum_{j=1}^{d}\kappa_{ij}h_{ji}=0 \quad \text{for all} \quad h.$$

Then the condition (7.4) is equivalent to $\kappa=0$, that is

$$H\hat{x}(\hat{x}'H\hat{x})^{-1}=K\hat{x}(\hat{x}'K\hat{x})^{-1} \quad \text{or} \quad cb=b(b'b)^{-1}(b'cb) \quad \text{where} \quad \hat{x}=K^{-\frac{1}{2}}b.$$

This means that cb is in the space spanned by b, and hence that the space $sp(b)$ is invariant under linear mapping c. To see this note that the matrix $(b'b)^{-1}(b'cb)$ is of dimension $r\times r$, then the columns of $b(b'b)^{-1}(b'cb)$ are linear combinations of those of b, and hence cb is in $sp(b)$. Using the property that any invariant subspace is spanned by a subset of eigenvectors, we have $sp(b)=sp(e_{i_1},\dots,e_{i_r})$ for some choice of non-collinear eigenvectors $e_{i_1},\dots,e_{i_r}$ of the matrix c. Since we have $\hat{x}=K^{-\frac{1}{2}}b$ we obtain $sp(\hat{x})=sp(K^{-\frac{1}{2}}(e_{i_1},\dots,e_{i_r}))$. In addition noting that $|\hat{x}'K\hat{x}|=|b'b|$ and $|\hat{x}'H\hat{x}|=|b'cb|=|b'b|\prod_{k=1}^{r}\delta_{i_k}$, we obtain $f(\hat{x})=\prod_{k=1}^{r}\delta_{i_k}$ which is clearly minimal if we choose $i_1,\dots,i_r$ among the set of the eigenvalues δ_{i_k} such that $\delta_{i_k}\leq\delta_r$. This complete the proof of Lemma 7.1. $\square$

In our framework we have to minimize the expression

$$|\beta'^{*}(S_{11}-S_{10}S_{00}^{-1}S_{01})\beta^{*}|/|\beta'^{*}S_{11}\beta^{*}|. \tag{7.5}$$

First we will proove that S_{11} is definite positive almost surely. Note that if S_{11} is not definite positive, then there exists $\iota_0\in\mathbb{R}^{d+1}$ such that

$$\iota_0'S_{11}\iota_0=\frac{1}{T}\sum_{t=1}^{T}\iota_0'R_{1t}R_{1t}'\iota_0=0$$

which entails $\iota_0'R_{1t}=0$ for $t=1,2,\dots,T$. From (2.2) we write

$$\iota_0'R_{1t}=\tilde{\iota}_0'K\epsilon_{t-1}+r_{t-1}, \tag{7.6}$$

where r_{t-1} is not correlated with ϵ_{t-1} and $\tilde{\iota}_0$ is given by the d first components of ι_0. Note that if the matrix K is not of full rank, then then there exists $\iota_0\neq 0$ such that one can predict $\iota_0'R_{1t}$ from it past values. It is easy to see that this is not consistent with the fact that Σ_ϵ is positive definite and then K is of full rank. From (7.6) we have $Var(\iota_0' R_{1t}) = Var(\tilde{\iota}_0' \epsilon_{t-1}) + Var(r_{t-1}) \geq \tilde{\iota}_0' \Sigma_\epsilon \tilde{\iota}_0 > 0$. Therefore $\iota_0' R_{1t} = 0$ is not almost surely equal to zero, and then S_{11} is almost surely positive definite. Note that using parallel arguments one can proove that S_{00} is almost surely definite positive. Now we will proove that the matrix $S_{11} - S_{10} S_{00}^{-1} S_{01}$ is definite positive. Consider the following matrix

$$\Theta = \begin{pmatrix} S_{00} & S_{01} \\ S_{10} & S_{11} \end{pmatrix}.$$

Similarly the assertion that Θ is not definite positive is equivalent to say that there exists $\iota = (\iota_1, \iota_2) \neq 0$ such that $\iota_1' R_{0t} + \iota_2' R_{1t} = 0$ where $\iota_1 \in \mathbb{R}^d$ and $\iota_2 \in \mathbb{R}^{d+1}$. Since we assumed that Σ_ϵ is positive definite, this not consistent with (2.9), and hence Θ is positive definite. Then writing

$$\Theta = \begin{pmatrix} I_d & 0 \\ S_{10} S_{00}^{-1} & I_{d+1} \end{pmatrix} \begin{pmatrix} S_{00} & 0 \\ 0 & S_{11} - S_{10} S_{00}^{-1} S_{01} \end{pmatrix} \begin{pmatrix} I_d & S_{00}^{-1} S_{01} \\ 0 & I_{d+1} \end{pmatrix} = \digamma \beth \digamma',$$

and noting that $\digamma$ is of full rank, it is easy to see that $\beth$ is definite positive. Then since all the principal minors of $\beth$ are positive implies that all the principal minors of $S_{11} - S_{10} S_{00}^{-1} S_{01}$ are positive, the result follow.

Thus from Lemma 7.1 the expression (7.5) is minimized by considering the eigenvectors corresponding to the r smallest solutions $\hat{\delta}_r \geq \cdots \geq \hat{\delta}_1 > 0$ of the eigenvalue problem

$$| \, (1-\delta) I_d - S_{11}^{-\frac{1}{2}} S_{10} S_{00}^{-1} S_{01} S_{11}^{-\frac{1}{2}} \, | = 0,$$

or equivalently the r largest solutions $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_r$ of the eigenvalue problem

$$| \, \lambda I - S_{11}^{-\frac{1}{2}} S_{10} S_{00}^{-1} S_{01} S_{11}^{-\frac{1}{2}} \, | = 0, \tag{7.7}$$

taking $\hat{\lambda}_i = 1 - \hat{\delta}_i$. The minimal value is therefore given by $\prod_{i=1}^r (1 - \hat{\lambda}_i)$ and we obtain

$$\hat{\beta}^* = S_{11}^{-\frac{1}{2}} (v_1, \ldots, v_r)$$

where $v_1, \ldots, v_r$ are the eigenvectors corresponding to the r largest solutions of (7.7).

Remark 7.1. In Lemma 7.1 note that if we have $\delta_{r+q} = \cdots = \delta_r$ for $q \in \{1, \ldots, d-r\}$, the space spanned by the various matrices of the form given in (7.2) is not unique. To see this suppose that $\delta_{r+1} = \delta_r$ then since the choice of the corresponding eigenvectors e_{r+1} and e_r in this case is not unique one can replace a given e_r by any vector of the eigenspace of δ_r. However in our case we show in the proof of Proposition 3.1 below that $\hat{\lambda}_{r_0}$ tends to a positive number and $\hat{\lambda}_{r_0+1}$ tends to zero at the rate T^{-1} as $T \to \infty$. Then there exists T_0 for which $\hat{\lambda}_{r_0} \neq \hat{\lambda}_{r_0+1}$ for all $T > T_0$. Thereby, under the null hypothesis, the uniqueness of the space spanned by $\hat{\beta}$ is ensured for a large enough T since it corresponds to the space spanned by the eigenspaces of the r_0 largest eigenvalues of (7.7) with $\hat{\lambda}_{r_0} \neq \hat{\lambda}_{r_0+1}$.

In our framework it is also important to see that we are estimating the space spanned by the columns of β_0^*. Therefore noting that when $\hat{\lambda}_{i_1} = \cdots = \hat{\lambda}_{i_q}$ with $i_1 \neq \cdots \neq i_q$ and $i_1, \ldots, i_q$ are smaller than r_0, the corresponding eigenvectors $v_{i_1}, \ldots, v_{i_q}$ are taken arbitrarily since the choice of these eigenvectors is not unique. Similarly we choose an arbitrarily order for the eigenvectors $v_1, \dots, v_{r_0}$ by taking $\hat{\beta}^* = (v_1, \dots, v_{r_0})$. In fact from the kind of normalization we use in section 2 these choices does not matter. Consider $\hat{\beta}_1^* = (\hat{\beta}_1', \hat{\tau}_1)'$ and $\hat{\beta}_2^* = (\hat{\beta}_2', \hat{\tau}_2)'$ such that $\hat{\beta}_1^* \neq \hat{\beta}_2^*$. Using a similar computations of (2.4), it is easy to see that $\hat{\beta}_{1c} = \hat{\beta}_{2c}$ and $\hat{\beta}_{1c}^* = \hat{\beta}_{2c}^*$.

Finally note that if we have $r_0 = 0$ we take $sp(\hat{\beta}^*) = \{0\}$ and therefore we do not need to apply Lemma 7.1 in this case.

In order to prove the results of our paper we have to state some intermediate asymptotic results. First we will state the following Lemma in which we use the mixing properties of the process (ϵ_t).

Lemma 7.2. *Under* **A2** *and* **A4** *we have*

$$\sup_{i,j} \sum_{h=-\infty}^{+\infty} \mid Cov(\epsilon_{m_1 t}\epsilon_{m_2 t-i}, \epsilon_{m_1' t-h}\epsilon_{m_2' t-j-h}) \mid < \infty,$$

where $m_1, m_2, m_1', m_2' \in \{1, \dots, d\}$.

Proof of Lemma 7.2. Note that without loss of generality, we can take $h \geq 0$ and $0 \leq i \leq j$. Then we write

$$\sum_{h=0}^{+\infty} \mid \mathrm{Cov}(\epsilon_{m_1 t}\epsilon_{m_2 t-i}, \epsilon_{m_1' t-h}\epsilon_{m_2' t-j-h}) \mid = a_1 + a_2.$$

where

$$a_1 = \sum_{h=0}^{i-1} \mid \mathrm{Cov}(\epsilon_{m_1 t}\epsilon_{m_2 t-i}, \epsilon_{m_1' t-h}\epsilon_{m_2' t-j-h}) \mid$$

and

$$a_2 = \sum_{h=i}^{+\infty} \mid \mathrm{Cov}(\epsilon_{m_1 t}\epsilon_{m_2 t-i}, \epsilon_{m_1' t-h}\epsilon_{m_2' t-j-h}) \mid .$$

Using the Davydov inequality (Davydov (1968)) and the Hölder inequality we have

$$a_2 \leq K_0 \parallel \epsilon_t \parallel_{4+2\nu}^4 \sum_{h=0}^{\infty} \{\alpha_\epsilon(h)\}^{\nu/(2+\nu)} < \infty,$$

where K_0 is an universal constant. To deal with the terms for $h < i$ we write

$$\begin{aligned} \mathrm{Cov}\left(\epsilon_{m_1 t}\epsilon_{m_2 t-i}, \epsilon_{m_1' t-h}\epsilon_{m_2' t-j-h}\right) &= \mathrm{Cov}\left(\epsilon_{m_1 t}\epsilon_{m_1' t-h}, \epsilon_{m_2 t-i}\epsilon_{m_2' t-j-h}\right) \\ &\quad + E\left\{\epsilon_{m_1 t}\epsilon_{m_1' t-h}\right\} E\left\{\epsilon_{m_2 t-i}\epsilon_{m_2' t-j-h}\right\} \\ &\quad - E\left\{\epsilon_{m_1 t}\epsilon_{m_2 t-i}\right\} E\left\{\epsilon_{m_1' t-h}\epsilon_{m_2' t-j-h}\right\} \quad (7.8) \end{aligned}$$

so that we have $a_1 \leq a_3 + a_4 + a_5$ where

$$a_3 = \sum_{h=0}^{i-1} \mathrm{Cov}\left(\epsilon_{m_1 t}\epsilon_{m_1' t-h}, \epsilon_{m_2 t-i}\epsilon_{m_2' t-j-h}\right),$$

$$a_4 = \sum_{h=0}^{i-1} E\left\{\epsilon_{m_1 t}\epsilon_{m_1' t-h}\right\} E\left\{\epsilon_{m_2 t-i}\epsilon_{m_2' t-j-h}\right\}$$

and

$$a_5 = \sum_{h=0}^{i-1} \left\{\epsilon_{m_1 t}\epsilon_{m_2 t-i}\right\} E\left\{\epsilon_{m_1' t-h}\epsilon_{m_2' t-j-h}\right\}.$$

Now it remains to check that the terms a_3, a_4 and a_5 are bounded. First note that

$$a_3 \leq K_0 \parallel \epsilon_t \parallel_{4+2\nu}^4 \sum_{h=0}^{i-1} \{\alpha_\epsilon(i-h)\}^{\nu/(2+\nu)} < \infty.$$

In addition we have using the Cauchy-Schwartz inequality and the Davydov inequality

$$\begin{aligned} a_4 &\leq \parallel \epsilon_t \parallel_2^2 \sum_{h=0}^{i-1} E\left\{\epsilon_{m_1 t}\epsilon_{m_1' t-h}\right\} \\ &\leq K_0 \parallel \epsilon_t \parallel_2^2 \parallel \epsilon_t \parallel_{2+\nu}^2 \sum_{h=0}^{\infty} \{\alpha_\epsilon(h)\}^{\nu/(2+\nu)}, \end{aligned}$$

and

$$\begin{aligned} a_5 &\leq \parallel \epsilon_t \parallel_2^2 iE\left\{\epsilon_{m_1 t}\epsilon_{m_2 t-i}\right\} \\ &\leq K_0 \parallel \epsilon_t \parallel_2^2 \parallel \epsilon_t \parallel_{2+\nu}^2 \sup_{i\geq 0} i\{\alpha_\epsilon(i)\}^{\nu/(2+\nu)}. \end{aligned}$$

Since $\sup_{i\geq 0} i\{\alpha_\epsilon(i)\}^{\nu/(2+\nu)} < \infty$, these two above expressions are bounded, and then the result follow. $\square$

Now define the linear process

$$V_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$$

where $\psi(z) = \sum_{i=0}^{\infty} \psi_i z^i$ is convergent for $\mid z \mid \leq 1+\delta$ for some $\delta > 0$. In the sequel we take $\sum_{i=1}^{j} \epsilon_t = 0$ when $j < 1$. The two following Lemmas provide us some useful results in our framework.

Lemma 7.3. *Under* **A2** *and* **A3** *we have*

$$T^{-\frac{1}{2}} \sum_{t=1}^{[Tu]} V_t \Rightarrow \psi(1)W(u), \tag{7.9}$$

$$T^{-1} \sum_{t=1}^{T} (\sum_{i=1}^{t-1} \epsilon_i) V_t' \Rightarrow \int_0^1 W(u)(dW)'\psi(1)' + \Sigma_\epsilon (\sum_{i=1}^{\infty} \psi_i)', \tag{7.10}$$

$$T^{-1} \sum_{t=1}^{T} (\sum_{i=1}^{t-1} \epsilon_i) V_{t-1}' \Rightarrow \int_0^1 W(u)(dW)'\psi(1)' + \Sigma_\epsilon \psi(1)', \tag{7.11}$$

$$T^{-\frac{3}{2}}\sum_{t=1}^{T} tV_t = O_p(1), \tag{7.12}$$

$$T^{-\frac{3}{2}}\sum_{t=1}^{T} tV_{t-1} = O_p(1), \tag{7.13}$$

where $W(u)$ is a brownian motion of variance Σ_ϵ.

Note that the result (7.9) is given in Phillips and Solo (1992) under the assumption that the process is a martingale difference, and similar results of (7.10), (7.10) and (7.12) can be found in Johansen (1995) in the iid case.

Proof of Lemma 7.3. To prove (7.9) we use the well known decomposition

$$\psi(z) = \psi(1) + (1-z)\psi^*(z)$$

where $\psi^*(z) = -\sum_{i=0}^{\infty}(\sum_{j=i+1}^{\infty}\psi_j)z^i$ and $V_t^* = \psi^*(L)\epsilon_t$, so that we obtain

$$V_t = \psi(1)\epsilon_t + \Delta V_t^*. \tag{7.14}$$

Then we write

$$\sum_{i=1}^{t} V_i = \psi(1)\sum_{i=1}^{t}\epsilon_t + V_t^* - V_0^*.$$

From the assumptions of our Lemma we have

$$\|V_t^*\|_{2+\nu+\eta} = \|\psi^*(L)\epsilon_t\|_{2+\nu+\eta} < \infty,$$

where L is the usual lag operator. Then using the Chebyshev inequality, we have

$$P\{\max_{1\le t\le T} \| V_t^* \| \ge \epsilon T^{\frac{1}{2}}\} \le \sum_{t=1}^{T} P\{\| V_t^* \| \ge \epsilon T^{\frac{1}{2}}\}$$

$$\le \epsilon^{-s} T^{\frac{2-s}{2}} E(\| V_1^* \|^s) \to 0, \tag{7.15}$$

for some $2 < s < 2+\nu+\eta$.

Noting that from the assumptions we made in our Lemma the process (ϵ_t) also verifies the mixing and moment conditions of **A3**, it follows from Herrndorf (1984, Corollary 1, p. 142) that

$$T^{-\frac{1}{2}}\sum_{t=1}^{[Tu]}\epsilon_t \Rightarrow W(u),$$

and then we obtain (7.9).

For the proof of (7.10) we write from (7.14)

$$T^{-1}\sum_{t=1}^{T}(\sum_{i=1}^{t-1}\epsilon_i)V_t' = T^{-1}\sum_{t=1}^{T}(\sum_{i=1}^{t-1}\epsilon_i)\epsilon_t'\psi(1)' + T^{-1}\sum_{t=1}^{T}(\sum_{i=1}^{t-1}\epsilon_i)\Delta V_t'^*.$$

Using the result in Phillips (1988) we obtain

$$T^{-1}\sum_{t=1}^{T}(\sum_{i=1}^{t-1}\epsilon_i)\epsilon_t'\psi(1)' \Rightarrow \int_0^1 W(u)(dW)'\psi(1)'.$$

In addition we have

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}(\sum_{i=1}^{t-1}\epsilon_i)\Delta V_t'^{*} &= T^{-1}(\sum_{t=1}^{T}\epsilon_t)V_T'^{*} - T^{-1}\sum_{t=1}^{T}\epsilon_t V_t'^{*} \\ &= T^{-\frac{1}{2}}(\sum_{t=1}^{T}\epsilon_t)T^{-\frac{1}{2}}V_T'^{*} - T^{-1}\sum_{t=1}^{T}\epsilon_t V_t'^{*}. \qquad (7.16) \end{aligned}$$

Using again the CLT given in Herrndorf (1984) and using (7.15), the first term in the right hand side of (7.16) converge to zero in probability by the Slutsky Lemma. For the second term using the fact that $\psi_0^* = -\sum_{i=1}^{\infty}\psi_i$ we obtain

$$T^{-1}\sum_{t=1}^{T}\epsilon_t V_t'^{*} \xrightarrow{P} -E(\epsilon_t V_t'^{*}) = \Sigma_\epsilon(\sum_{i=1}^{\infty}\psi_i)'.$$

Then the result (7.10) follow. For the proof of (7.11) we write

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}(\sum_{i=1}^{t-1}\epsilon_i)V_{t-1}' &= T^{-1}\sum_{t=1}^{T}(\sum_{i=1}^{t-1}\epsilon_i)\epsilon_{t-1}'\psi(1)' + T^{-1}\sum_{t=1}^{T}(\sum_{i=1}^{t-1}\epsilon_i)\Delta V_{t-1}'^{*} \\ &= T^{-1}\sum_{t=2}^{T}\epsilon_{t-1}\epsilon_{t-1}'\psi(1)' + T^{-1}\sum_{t=1}^{T}(\sum_{i=1}^{t-2}\epsilon_i)\epsilon_{t-1}'\psi(1)' \\ &\quad + T^{-1}\sum_{t=1}^{T}(\sum_{i=1}^{t-1}\epsilon_i)\Delta V_{t-1}'^{*}. \end{aligned}$$

Using a similar decomposition of (7.16) we have

$$T^{-1}\sum_{t=1}^{T}(\sum_{i=1}^{t-1}\epsilon_i)\Delta V_{t-1}'^{*} \xrightarrow{P} -E(\epsilon_t V_{t-1}'^{*}) = 0.$$

Noting that $T^{-1}\sum_{t=1}^{T}\epsilon_{t-1}\epsilon_{t-1}' \xrightarrow{P} \Sigma_\epsilon$, it is easy to see that we obtain (7.11) using similar arguments of the proof of (7.10).

For the proof of (7.12) note that $\| \frac{t}{T}V_t \| \leq \| V_t \|$ and then the statement (7.12) follows from (7.9). Finally for the proof of (7.13), noting that from (7.9) it can be shown that $T^{-\frac{1}{2}}\sum_{t=1}^{T}V_{t-1} = O_p(1)$, the result (7.13) follow in a similar way of (7.12). $\square$

Lemma 7.4. *Under* **A2** *and* **A4** *we have*

$$T^{-\frac{1}{2}}\sum_{t=1}^{T}vec(\epsilon_t V_{t-1}') \Rightarrow \mathcal{N}(0,\Xi),$$

where the matrix Ξ is of the form

$$\Xi = \sum_{h=-\infty}^{\infty} E\left\{V_{t-1} \otimes \epsilon_t\right\}\left\{V_{t-h-1} \otimes \epsilon_{t-h}\right\}'.$$

If we assume that the error process is iid, we obtain $\Xi = \Sigma_V \otimes \Sigma_\epsilon$ where $\Sigma_V = E(V_t V_t')$.

Proof of Lemma 7.4. Let us define $u_t = vec(\epsilon_t V_{t-1}') = \sum_{i=0}^{\infty} vec(\epsilon_t \epsilon_{t-i-1}' \psi_i')$. We also define $u_{q,t} = \sum_{i=0}^{q} vec(\epsilon_t \epsilon_{t-i-1}' \psi_i')$, where $q \sim T^\gamma$ for some $\gamma \in]0,1[$. With these notations we write

$$u_t = u_{q,t} + e_{q,t} \quad \text{where} \quad e_{q,t} = \sum_{i=q+1}^{\infty} vec(\epsilon_t \epsilon_{t-i-1}' \psi_i').$$

From Lemma 7.2 and using the Chebyshev inequality and the fact that the coefficients of the matrices ψ_i decay exponentially it can be shown that $T^{-\frac{1}{2}} \sum_{t=1}^{T} e_{q,t} = o_p(1)$. Then we can deduce that $T^{-\frac{1}{2}} \sum_{t=1}^{T} u_t$ and $T^{-\frac{1}{2}} \sum_{t=1}^{T} u_{q,t}$ has the same asymptotic behaviour.

From the expression of $u_{q,t}$ we obviously have $\|u_{q,t}\|_{4+2\nu} < \infty$. In addition we have $\alpha_{u_q}(h-q) \le \alpha_\epsilon(h)$, so that $\sum_{h=0}^{\infty} \{\alpha_{u_q}(h)\}^{\nu/(2+\nu)} < \infty$. Noting that

$$u_{q,t} = \sum_{i=0}^{q} vec(\epsilon_t \epsilon_{t-i-1}' \psi_i') = \sum_{i=0}^{q} (\psi_i \otimes I_d)(\epsilon_{t-i-1} \otimes \epsilon_t),$$

we write using the Lebesgue theorem and the stationarity of $u_{q,t}$

$$\begin{aligned} & \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathrm{cov}(u_{q,t}, u_{q,s}) = \lim_{T\to\infty} \frac{1}{T} \sum_{|h|<T} (T-\mid h \mid) \mathrm{cov}(u_{q,t}, u_{q,t-h}) \\ = \; & \lim_{T\to\infty} \frac{1}{T} \sum_{i,j=1}^{q} \sum_{|h|<T} (T-\mid h \mid)(\psi_i \otimes I_d) \mathrm{cov}\{(\epsilon_{t-i-1} \otimes \epsilon_t), (\epsilon_{t-i-h-1} \otimes \epsilon_{t-h})'\} \\ & (\psi_i' \otimes I_d) = \sum_{i,j=1}^{\infty} \sum_{h=-\infty}^{\infty} (\psi_i \otimes I_d) \mathrm{cov}\{(\epsilon_{t-i-1} \otimes \epsilon_t), (\epsilon_{t-i-h-1} \otimes \epsilon_{t-h})'\} \\ & (\psi_i' \otimes I_d). \end{aligned}$$

The existence of this last sum is ensured by Lemma 7.2 and using the fact that the coefficients of the matrices ψ_i decay exponentially. Then from the CLT given in Herrndorf (1984), $T^{-\frac{1}{2}} \sum_{t=1}^{T} u_{q,t}$ is normally distributed with mean zero. We obtain the expression of Ξ writing

$$u_t = vec(\epsilon_t V_{t-1}') = (V_{t-1} \otimes I_d)\epsilon_t = V_{t-1} \otimes \epsilon_t,$$

and

$$\begin{aligned} & \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathrm{cov}(u_t, u_s) = \lim_{T\to\infty} \frac{1}{T} \sum_{|h|<T} (T-\mid h \mid) \mathrm{cov}(u_t, u_{t-h}) \\ = \; & \sum_{h=-\infty}^{\infty} \mathrm{cov}\left(u_t, u_{t-h}\right) = \sum_{h=-\infty}^{\infty} E\left\{V_{t-1} \otimes \epsilon_t\right\}\left\{V_{t-h-1} \otimes \epsilon_{t-h}\right\}'. \end{aligned}$$

This complete the proof of our Lemma. □

The following Lemmas are equivalent to Lemmas 10.2 and 10.3 in Johansen (1995). Recall that $\bar{\beta}_{0\perp} = \beta_{0\perp}(\beta'_{0\perp}\beta_{0\perp})^{-1}$.

Lemma 7.5. *Under* **A1**, **A2** *and* **A3**, *the process* Z_{1t} *satisfies*

$$T^{-\frac{1}{2}} C'_T (Z_{1[Tu]} - \bar{Z}_1) \Rightarrow G(u) \tag{7.17}$$

where

$$G(u) = \begin{pmatrix} \bar{\beta}'_{0\perp} C(W(u) - \bar{W}) \\ -u + \frac{1}{2} \end{pmatrix}, \quad \bar{Z}_1 = T^{-1}\sum_{t=1}^{T} Z_{1t}, \bar{W} = \int_0^1 W(u)du,$$

and

$$C_T = \begin{pmatrix} \bar{\beta}_{0\perp} & 0 \\ \rho'_{o1}\bar{\beta}_{0\perp} & T^{-\frac{1}{2}} \end{pmatrix}.$$

Proof of Lemma 7.5. From (2.2) we have

$$\begin{aligned} T^{-\frac{1}{2}}(\bar{\beta}'_{0\perp}, \bar{\beta}'_{0\perp}\rho_{o1}) Z_{1[Tu]} &= T^{-\frac{1}{2}}\bar{\beta}'_{0\perp} C \sum_{i=1}^{[T(u-\frac{1}{T})]} \epsilon_i + T^{-\frac{1}{2}}\bar{\beta}'_{0\perp} Y_{[Tu]-1} \\ &\quad + T^{-\frac{1}{2}}\bar{\beta}'_{0\perp}(\rho_{o1} + \rho_{o0} + A). \end{aligned} \tag{7.18}$$

It can be easily shown that the second term on the right hand side tends to zero in probability using the Chebyshev inequality. In addition the third term does not depends on time and vanishes by the factor $T^{-\frac{1}{2}}$. From **A3** and using the central limit theorem given by Herrndorf (1984) it follows

$$T^{-\frac{1}{2}}\bar{\beta}'_{0\perp} C \sum_{i=1}^{[Tu]} \epsilon_i \Rightarrow \bar{\beta}'_{0\perp} C W(u).$$

Finally considering the continuous mapping $x \longrightarrow \int_0^1 x(u)du$, we obtain from the continuous mapping theorem

$$T^{-\frac{1}{2}}(\bar{\beta}'_{0\perp}, \bar{\beta}'_{0\perp}\rho_{o1})\bar{Z}_{1t} = T^{-1}\sum_{t=1}^{T} T^{-\frac{1}{2}}(\bar{\beta}'_{0\perp}, \bar{\beta}'_{0\perp}\rho_{o1}) Z_{1t} \Rightarrow \bar{\beta}'_{0\perp} C\bar{W}. \tag{7.19}$$

The asymptotic behaviour of the last component can be obtained noting that

$$\lim_{T\to\infty} \frac{-[Tu]+1}{T} = \lim_{T\to\infty} \frac{-[Tu]+Tu}{T} + \frac{1}{T} - u = -u,$$

and

$$\lim_{T\to\infty} -T^{-2}\sum_{t=1}^{T}(-t+1) = \lim_{T\to\infty} \frac{T(T+1)}{2T^2} - \frac{T}{T^2} = \frac{1}{2}.$$

□

The results of Lemmas 7.6, 7.7 and the proof of Proposition 4.1 are not modified by the choice of a normalization. Then we will consider for these results any β_0^*.

Lemma 7.6. *Under* **A1**, **A2** *and* **A3**, *the residuals* R_{1t} *satisfy*

$$T^{-1}C_T'S_{11}C_T \Rightarrow \int_0^1 GG'du \tag{7.20}$$

$$C_T'(S_{10} - S_{11}\beta_0^*\alpha_0') \Rightarrow \int_0^1 G(dW)' \tag{7.21}$$

$$C_T'S_{11}\beta_0^* = O_p(1) \tag{7.22}$$

$$C_T'S_{10} = O_p(1). \tag{7.23}$$

Proof of Lemma 7.6. First note that from (2.2) we have

$$Z_{0t} = \Delta X_t = C\epsilon_t + \rho_{o1} + \Delta Y_t, \tag{7.24}$$

and since $\beta_0'\rho_{o0} = \tau_0$ we write

$$\beta_0'^* Z_{1t} = \beta_0' X_{t-1} - \tau_0(t-1) = \beta_0'\rho_{o0} + \beta_0' Y_{t-1}. \tag{7.25}$$

Since the process Y_t is a stationary linear process, then it is easy to see that the centered processes $\beta_0'^*(Z_{1t} - \bar{Z}_1)$ and $(Z_{0t} - \bar{Z}_0)$ are $I(0)$ and that these processes can also be written as linear processes. Then we can use the results in Lemma 7.3 when needed. Define the centered stationary process $(\tilde{Z}_{2t} - \bar{\tilde{Z}}_2)$ where $\tilde{Z}_{2t} = (Z_{0t-1}', \dots, Z_{0t-p+1}')'$, and let us introduce the following notations

$$N_{11} = T^{-1}\sum_{t=1}^{T}(Z_{1t} - \bar{Z}_1)(Z_{1t} - \bar{Z}_1)',$$

$$N_{22} = T^{-1}\sum_{t=1}^{T}(\tilde{Z}_{2t} - \bar{\tilde{Z}}_2)(\tilde{Z}_{2t} - \bar{\tilde{Z}}_2)',$$

$$N_{12} = T^{-1}\sum_{t=1}^{T}(Z_{1t} - \bar{Z}_1)(\tilde{Z}_{2t} - \bar{\tilde{Z}}_2)',$$

$$N_{10} = T^{-1}\sum_{t=1}^{T}(Z_{1t} - \bar{Z}_1)(Z_{0t} - \bar{Z}_0)',$$

and

$$N_{20} = T^{-1}\sum_{t=1}^{T}(\tilde{Z}_{2t} - \bar{\tilde{Z}}_2)(Z_{0t} - \bar{Z}_0)'.$$

To prove (7.20) note that since we have $\Psi_0(Z_{2t}-\bar{Z}_2)=\tilde{\Psi}_0(\tilde{Z}_{2t}-\bar{\tilde{Z}}_2)$ where $\tilde{\Psi}=(\Gamma_1,\ldots,\Gamma_{p-1})$, we write from (2.7)

$$Z_{0t}-\bar{Z}_0=\alpha_0\beta_0^{\prime *}(Z_{1t}-\bar{Z}_1)+\tilde{\Psi}_0(\tilde{Z}_{2t}-\bar{\tilde{Z}}_2)+\epsilon_t.$$

Since we defined the $R_{1t}{}'s$ as the residuals of the regression of Z_{1t} on Z_{2t} we have

$$T^{-1}C_T'S_{11}C_T=T^{-1}C_T'N_{11}C_T-T^{-1}C_T'N_{12}N_{22}^{-1}N_{21}C_T. \tag{7.26}$$

Using (7.18) the d first rows of $C_T'N_{12}$ are of the form

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{T}(\bar{\beta}_{0\perp}',\bar{\beta}_{0\perp}'\rho_{o1})(Z_{1t}-\bar{Z}_1)(\tilde{Z}_{2t}-\bar{\tilde{Z}}_2)'=\\
&T^{-1}\bar{\beta}_{0\perp}'C\sum_{t=1}^{T}(\sum_{i=1}^{t-1}\epsilon_i)(\tilde{Z}_{2t}-\bar{\tilde{Z}}_2)'+T^{-1}\bar{\beta}_{0\perp}'\sum_{t=1}^{T}Y_{t-1}(\tilde{Z}_{2t}-\bar{\tilde{Z}}_2)'\\
&+T^{-1}\sum_{t=1}^{T}\bar{\beta}_{0\perp}'(\rho_{o1}+\rho_{o0}+A)(\tilde{Z}_{2t}-\bar{\tilde{Z}}_2)'\\
&-T^{-1}\sum_{t=1}^{T}(\bar{\beta}_{0\perp}',\bar{\beta}_{0\perp}'\rho_{o1})\bar{Z}_1(\tilde{Z}_{2t}-\bar{\tilde{Z}}_2)'.
\end{aligned}\tag{7.27}$$

Note that from the expression of $(\tilde{Z}_{2t}-\bar{\tilde{Z}}_2)$ it is easy to see that this process is of the form

$$\tilde{Z}_{2t}-\bar{\tilde{Z}}_2=\sum_{i=0}^{\infty}\dot{\psi}_i(\epsilon_{t-i-1}',\ldots,\epsilon_{t-i-p+1}')'.$$

Then using (7.11) the first term on the right hand side of (7.27) is normalized to converge. The processes in the second and third terms in (7.27) are stationary ergodic, and then using the ergodic theorem it is easy to see that these terms are normalized to converge. Finally note that since $\bar{Z}_1$ does not depend on t the last term can be written as $\{T^{-\frac{1}{2}}(\bar{\beta}_{0\perp}',\bar{\beta}_{0\perp}'\rho_{o1})\bar{Z}_1\}\{T^{-\frac{1}{2}}\sum_{t=1}^{T}(\tilde{Z}_{2t}-\bar{\tilde{Z}}_2)'\}$. From (7.19) the term $T^{-\frac{1}{2}}(\bar{\beta}_{0\perp}',\bar{\beta}_{0\perp}'\rho_{o1})\bar{Z}_1$ converge weakly, and using (7.9) the term $\{T^{-\frac{1}{2}}\sum_{t=1}^{T}(\tilde{Z}_{2t}-\bar{\tilde{Z}}_2)'\}$ also converge. Moreover the last row of $C_T'N_{12}$ is of the form

$$\begin{aligned}
&T^{-\frac{3}{2}}\sum_{t=1}^{T}\{-t+1-\sum_{t=1}^{T}\frac{-t+1}{T}\}(\tilde{Z}_{2t}-\bar{\tilde{Z}}_2)'=\frac{1}{2}T^{-\frac{1}{2}}\sum_{t=1}^{T}(\tilde{Z}_{2t}-\bar{\tilde{Z}}_2)'\\
&+\frac{1}{2}T^{-\frac{3}{2}}\sum_{t=1}^{T}(\tilde{Z}_{2t}-\bar{\tilde{Z}}_2)'-T^{-\frac{3}{2}}\sum_{t=1}^{T}t(\tilde{Z}_{2t}-\bar{\tilde{Z}}_2)'.
\end{aligned}\tag{7.28}$$

From (7.9) the first and the second term in the right hand side of (7.28) converge, whereas the third term converge from (7.13). Thus we can conclude that the matrix $C_T'N_{12}$ is normalized to converge. In addition using the ergodic theorem for the strictly stationary process $(\tilde{Z}_{2t-1}-\bar{\tilde{Z}}_2)$ the term N_{22} converges to its population value. Therefore the second term in the right hand side of (7.26) tends to zero by the factor

T^{-1}. On the other hand considering the continuous mapping $x \longrightarrow \int_0^1 x(u)x(u)'du$, it follow from the continuous mapping theorem and Lemma 7.5 that

$$T^{-1}C_T'N_{11}C_T \Rightarrow \int_0^1 GG'du,$$

which completes the proof of (7.20).

Similarly for the proof of (7.22) we write

$$C_T'S_{11}\beta_0^* = C_T'N_{11}\beta_0^* - C_T'N_{12}N_{22}^{-1}N_{21}\beta_0^*. \tag{7.29}$$

First note that the rows of the matrix $C_T'N_{11}\beta_0^*$ can be written in the same way of those of the matrix $C_T'N_{12}$ replacing only $\tilde{Z}_{2t} - \bar{\tilde{Z}}_2$ by $\beta_0'^*(Z_{1t} - \bar{Z}_1)$. Since the process $\beta_0'^*(Z_{1t} - \bar{Z}_1)$ is also stationary and can be written as a linear process, then considering the arguments we used for the matrix $C_T'N_{12}$ one can show that the matrix $C_T'N_{11}\beta_0^*$ is normalized to converge. Finally noting that the processes $(\tilde{Z}_{2t} - \bar{\tilde{Z}}_2)$ and $\beta_0'^*(Z_{1t} - \bar{Z}_1)$ are stationary ergodic the term $N_{21}\beta_0^*$ converges using the Cauchy-Schwarz inequality and the ergodic theorem. Then since the terms in the right hand side of (7.29) are convergent we obtain the result (7.22).

For the proof of (7.23) we write

$$C_T'S_{10} = C_T'N_{10} - C_T'N_{12}N_{22}^{-1}N_{20}. \tag{7.30}$$

Similarly we can show that the matrix $C_T'N_{10}$ converge using the same arguments considered for the matrix $C_T'N_{12}$ and replacing $\tilde{Z}_{2t} - \bar{\tilde{Z}}_2$ by $Z_{0t} - \bar{Z}_0$. However note that since from (7.24) the term $Z_{0t} - \bar{Z}_0$ is of the form

$$Z_{0t} - \bar{Z}_0 = \sum_{i=0}^{\infty} \ddot{\psi}_i \epsilon_{t-i},$$

we shall use in this case relations (7.10) and (7.12) to conclude. In addition since the process $(Z_{0t} - \bar{Z}_0)$ is stationary a process, then the matrix N_{20} converge. Therefore the matrices in the right hand side of (7.30) are all normalized to converge and the result (7.23) follows.

To prove (7.21) note that from (2.9) we have

$$C_T'(S_{10} - S_{11}\beta_0^*\alpha_0') = C_T'N_\epsilon = C_T'N_{1\epsilon} - C_T'N_{12}N_{22}^{-1}N_{2\epsilon} \tag{7.31}$$

where

$$N_\epsilon = T^{-1}\sum_{t=1}^{T} R_{1t}\epsilon_t', \quad N_{1\epsilon} = T^{-1}\sum_{t=1}^{T}(Z_{1t} - \bar{Z}_1)\epsilon_t',$$

$$\text{and} \quad N_{2\epsilon} = T^{-1}\sum_{t=1}^{T}(\tilde{Z}_{2t} - \bar{\tilde{Z}}_2)\epsilon_t'.$$

From the ergodic theorem and since $\tilde{Z}_{2t}$ and ϵ_t are uncorrelated, the term $N_{2\epsilon}$ tends to zero in probability. Then the second term in the right hand side of (7.31) tend to zero. Finally using Lemma (7.5) and the continuous mapping theorem we write

$$C_T'N_{1\epsilon} \Rightarrow \int_0^1 G(dW)'. \tag{7.32}$$

This complete the proof of our Lemma. □

Now let us define the following matrices

$$\Sigma_{ij} = \Lambda_{ij} - \Lambda_{i2}\Lambda_{22}^{-1}\Lambda_{2j}$$

for $i, j = 0, 2, \beta$ and where the matrices Λ_{ij} are defined by,

$$\Lambda_{\beta\beta} = Var(\beta_0'^{*} Z_{1t}), \quad \Lambda_{00} = Var(Z_{0t}), \quad \Lambda_{22} = Var(\tilde{Z}_{2t}),$$

$$\Lambda_{\beta 0} = Cov(\beta_0'^{*} Z_{1t}, Z_{0t}), \quad \Lambda_{\beta 2} = Cov(\beta_0'^{*} Z_{1t}, \tilde{Z}_{2t}) \quad \text{and} \quad \Lambda_{20} = Cov(\tilde{Z}_{2t}, Z_{0t}).$$

Note that when β_0^* is normalized by the matrix c, we have $\Sigma_{\beta\beta} = \Sigma_c$, where Σ_c is defined in Section 3. The following Lemma provides us a result on the asymptotic behaviour of the matrices S_{11}, S_{00} and S_{10} in terms of the above defined matrices.

Lemma 7.7. *Under* **A1**, **A2** *and* **A3** *we have*

$$\beta_0'^{*} S_{11} \beta_0^{*} \xrightarrow{P} \Sigma_{\beta\beta} \tag{7.33}$$

$$\beta_0'^{*} S_{10} \xrightarrow{P} \Sigma_{\beta 0} \tag{7.34}$$

$$S_{00} \xrightarrow{P} \Sigma_{00} \tag{7.35}$$

where the matrices Σ_{00}, $\Sigma_{\beta 0}$ *and* $\Sigma_{\beta\beta}$ *verify*

$$\Sigma_{00} = \alpha_0 \Sigma_{\beta 0} + \Sigma_{\epsilon}, \quad \Sigma_{0\beta} = \alpha_0 \Sigma_{\beta\beta}, \tag{7.36}$$

and

$$\Sigma_{\epsilon} = \Sigma_{00} - \alpha_0 \Sigma_{\beta\beta} \alpha_0'. \tag{7.37}$$

Moreover we have

$$\Sigma_{00}^{-1} - \Sigma_{00}^{-1}\alpha_0(\alpha_0'\Sigma_{00}^{-1}\alpha_0)^{-1}\alpha_0'\Sigma_{00}^{-1} = \alpha_{0\perp}(\alpha_{0\perp}'\Sigma_{\epsilon}\alpha_{0\perp})^{-1}\alpha_{0\perp}'. \tag{7.38}$$

Proof of Lemma 7.7. Similarly to (7.26) we write

$$\beta_0'^{*} S_{11} \beta_0^{*} = \beta_0'^{*} N_{11} \beta_0^{*} - \beta_0'^{*} N_{12} N_{22}^{-1} N_{21} \beta_0^{*}.$$

On the other hand from (7.24) and (7.25) the processes $(\beta_0'^{*} Z_{1t})$, (Z_{0t}) and $(\tilde{Z}_{2t})$ are stationary ergodic since (Y_t) is stationary ergodic. Thus we have from the ergodic theorem

$$\beta_0'^{*} N_{11} \beta_0^{*} \xrightarrow{P} \Lambda_{\beta\beta}, \quad \beta_0'^{*} N_{12} \xrightarrow{P} \Lambda_{\beta 2}, \quad \text{and} \quad N_{22} \xrightarrow{P} \Lambda_{22},$$

which gives us the result (7.33). The proof of (7.34) and (7.35) are similar.
For the proof of the relations in (7.36), multiplying the expression (2.7) by $(Z_{0t} - Z_0)'$ and $(Z_{1t} - \bar{Z}_1)'\beta_0^*$ on the right, we have

$$\Lambda_{00} = \alpha_0 \Lambda_{\beta 0} + \tilde{\Psi}\Lambda_{20} + \Sigma_{\epsilon} \quad \text{and} \quad \Lambda_{0\beta} = \alpha_0 \Lambda_{\beta\beta} + \tilde{\Psi}\Lambda_{2\beta} \tag{7.39}$$

since we assumed that the error process (ϵ_t) is uncorrelated. Using again the expression (2.7) we write

$$\tilde{\Psi} = \Lambda_{02}\Lambda_{22}^{-1} - \alpha_0\Lambda_{\beta 2}\Lambda_{22}^{-1}. \tag{7.40}$$

Inserting (7.40) in the expressions in (7.39) we obtain the desired results. The expression (7.37) is a straightforward consequence of (7.36). The relation in (7.38) can be obtained using the following projection identity

$$I_d = \Sigma_{00}^{-1}\alpha_0(\alpha_0'\Sigma_{00}^{-1}\alpha_0)^{-1}\alpha_0' + \alpha_{0\perp}(\alpha_{0\perp}'\Sigma_{00}\alpha_{0\perp})^{-1}\alpha_{0\perp}'\Sigma_{00},$$

and noting that from (7.36) $\alpha_{0\perp}\Sigma_{00} = \alpha_{0\perp}\Sigma_\epsilon$. □

For the proof of Propositions 3.1 and 4.1 note that the solutions $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_{d+1}$ of the equation (2.11) are the same of those of the following eigenvalue problem

$$\mid \lambda S_{11} - S_{10}S_{00}^{-1}S_{01} \mid = 0. \tag{7.41}$$

The eigenvectors e_i of (7.41) which verify

$$S_{10}S_{00}^{-1}S_{01}e_i = \hat{\lambda}_i S_{11}e_i,$$

are such that $e_i = S_{11}^{-\frac{1}{2}}v_i$. Using this notation we write $\hat{\beta}^* = (e_1, \ldots, e_{r_0})$. Note also that since the matrix $S_{10}S_{00}^{-1}S_{01}$ is of dimension $d+1$ but has rank d, then $\hat{\lambda}_{d+1} = 0$.

Proof of Proposition 3.1. We first show that the roots $\hat{\lambda}_{r_0+1}, \ldots, \hat{\lambda}_d$ of (7.41) decrease at the rate T^{-1}. Let the matrix $A_T = (\beta_0^*, T^{-\frac{1}{2}}\ C_T)$. Multiplying (7.41) by A_T' and A_T, and noting that the matrix A_T is an invertible matrix, the equation

$$\mid A_T'(\lambda S_{11} - S_{10}S_{00}^{-1}S_{01})A_T \mid = \begin{vmatrix} \beta_0'^* S(\lambda)\beta_0^* & T^{-\frac{1}{2}}\beta_0'^* S(\lambda)C_T \\ T^{-\frac{1}{2}}C_T'S(\lambda)\beta_0^* & T^{-1}C_T'S(\lambda)C_T \end{vmatrix} = 0, \tag{7.42}$$

has the same eigenvalues as (7.41). From Lemmas 7.6 and 7.7 and since the solutions of (7.41) are continuous functions of the coefficient of the matrices S_{11}, S_{10}, S_{00}, and S_{01}, it follows that

$$\begin{aligned} \mid A_T'(\lambda S_{11} - S_{10}S_{00}^{-1}S_{01})A_T \mid &\Rightarrow \begin{vmatrix} \lambda\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} & 0 \\ 0 & \lambda\int_0^1 GG'du \end{vmatrix} \\ &= \mid \lambda\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} \mid\mid \lambda\int_0^1 GG'du \mid . \end{aligned}$$

Therefore there is r_0 roots of the equation (7.42) which converge to the r_0 positive roots given by the equation $\mid \lambda\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} \mid = 0$, and $d - r_0 + 1$ roots of (7.42) which converge to the $d - r_0 + 1$ zero roots given by the solutions of the equation $\mid \lambda\int_0^1 GG'du \mid = 0$. Defining $S(\lambda) = \lambda S_{11} - S_{10}S_{00}^{-1}S_{01}$ and using the relation

$$\begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} = \mid A_{11} \mid\mid A_{22} - A_{21}A_{11}^{-1}A_{12} \mid \tag{7.43}$$

in (7.42) for λ such that $\mid \beta_0'^* S(\lambda)\beta_0^* \mid \neq 0$, we write

$$\begin{vmatrix} \beta_0'^* S(\lambda)\beta_0^* & T^{-\frac{1}{2}}\beta_0'^* S(\lambda) C_T \\ T^{-\frac{1}{2}} C_T' S(\lambda)\beta_0^* & T^{-1} C_T' S(\lambda) C_T \end{vmatrix} = \mid \beta_0'^* S(\lambda)\beta_0^* \mid\mid \lambda\{T^{-1}C_T' S_{11} C_T\}$$
$$-T^{-1}\{C_T' S_{10} S_{00}^{-1} S_{01} C_T + \beta_0'^* S(\lambda) C_T (\beta_0'^* S(\lambda)\beta_0^*)^{-1} C_T' S(\lambda)\beta_0^*\} \mid .$$

It is seen that the roots which correspond to the eigenvalue problem $\mid \beta_0'^* S(\lambda)\beta_0^* \mid = 0$ do not converge to zero and have the same limit of the r greatest roots of (7.42). Then for a large T, the roots $\hat\lambda_{r_0+1}, \ldots, \hat\lambda_{d+1}$ cannot be in the set of the r_0 roots of $\mid \beta_0'^* S(\lambda)\beta_0^* \mid = 0$. It follows that $\hat\lambda_{r_0+1}, \ldots, \hat\lambda_{d+1}$ are solutions of the following equation

$$\begin{aligned} &\mid \lambda\{T^{-1}C_T' S_{11} C_T\} - T^{-1}\{C_T' S_{10} S_{00}^{-1} S_{01} C_T \\ &+\beta_0'^* S(\lambda) C_T (\beta_0'^* S(\lambda)\beta_0^*)^{-1} C_T' S(\lambda)\beta_0^*\} \mid = 0. \end{aligned} \tag{7.44}$$

Considering the roots $\hat\lambda_{r_0+1}, \ldots, \hat\lambda_d$ which converge to zero, and using the results of Lemmas 7.6 and 7.7 the terms into brackets in (7.44) are normalized to converge, then it is seen that the roots $\hat\lambda_{r_0+1}, \ldots, \hat\lambda_d$ of (7.41) decrease at the rate T^{-1}.

Now we will establish the asymptotic behaviour of the likelihood ratio test statistic. Using again the relation (7.43) we write

$$\begin{aligned} &\mid (\beta_0^*, C_T)' S(\lambda)(\beta_0^*, C_T) \mid = \begin{vmatrix} \beta_0'^* S(\lambda)\beta_0^* & \beta_0'^* S(\lambda) C_T \\ C_T' S(\lambda)\beta_0^* & C_T' S(\lambda) C_T \end{vmatrix} \\ &= \mid \beta_0'^* S(\lambda)\beta_0^* \mid\mid C_T' S(\lambda) C_T \\ &- C_T' S(\lambda)\beta_0^* (\beta_0'^* S(\lambda)\beta_0^*)^{-1} \beta_0'^* S(\lambda) C_T \mid = 0. \end{aligned} \tag{7.45}$$

For the rest of the proof we will focus on the second term of the right hand side of (7.45) and only consider the $d - r_0 + 1$ smallest roots $\hat\lambda_{r_0+1}, \ldots, \hat\lambda_{d+1}$. Noting that from the first part of the proof the roots $\hat\lambda_{r_0+1}, \ldots, \hat\lambda_d$ decrease at the rate T^{-1}, we therefore define $\eta = T\lambda$ where η is real. From Lemma 7.7 and using (7.22) we have

$$\beta_0'^* S(\lambda)\beta_0^* = \eta T^{-1}\beta_0'^* S_{11}\beta_0^* - \beta_0'^* S_{10} S_{00}^{-1} S_{01}\beta_0^* = -\Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} + o_p(1), \tag{7.46}$$

$$\begin{aligned} C_T' S(\lambda)\beta_0^* &= \eta T^{-1} C_T' S_{11}\beta_0^* - C_T' S_{10} S_{00}^{-1} S_{01}\beta_0^* \\ &= -C_T' S_{10}\Sigma_{00}^{-1}\Sigma_{0\beta} + o_p(1). \end{aligned} \tag{7.47}$$

Then inserting (7.46) and (7.47) into the second factor in (7.45) we obtain

$$\begin{aligned} &C_T' S(\lambda) C_T - C_T' S(\lambda)\beta_0^* (\beta_0'^* S(\lambda)\beta_0^*)^{-1}\beta_0'^* S(\lambda) C_T \\ &= \eta T^{-1} C_T' S_{11} C_T - C_T' S_{10}\Sigma_{00}^{-1} S_{01} C_T \\ &+ C_T' S_{10}\Sigma_{00}^{-1}\Sigma_{0\beta}(\Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta})^{-1}\Sigma_{\beta 0}\Sigma_{00}^{-1} S_{01} C_T + o_p(1) \\ &= \eta T^{-1} C_T' S_{11} C_T - C_T' S_{10} D S_{01} C_T + o_p(1), \end{aligned} \tag{7.48}$$

where from (7.36) and (7.38) the matrix D is given by

$$\begin{aligned} D &= \Sigma_{00}^{-1} - \Sigma_{00}^{-1}\Sigma_{0\beta}(\Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta})^{-1}\Sigma_{\beta 0}\Sigma_{00}^{-1} \\ &= \Sigma_{00}^{-1} - \Sigma_{00}^{-1}\alpha_0\Sigma_{\beta\beta}(\Sigma_{\beta\beta}\alpha_0'\Sigma_{00}^{-1}\alpha_0\Sigma_{\beta\beta})^{-1}\Sigma_{\beta\beta}\alpha_0'\Sigma_{00}^{-1} \\ &= \Sigma_{00}^{-1} - \Sigma_{00}^{-1}\alpha_0(\alpha_0'\Sigma_{00}^{-1}\alpha_0)^{-1}\alpha_0'\Sigma_{00}^{-1} \\ &= \alpha_{0\perp}(\alpha_{0\perp}'\Sigma_\epsilon\alpha_{0\perp})^{-1}\alpha_{0\perp}' = \alpha_{0\perp}(Var(\alpha_{0\perp}' W))^{-1}\alpha_{0\perp}'. \end{aligned} \tag{7.49}$$

From Lemma 7.6 we have

$$C_T' S_{10}\alpha_{0\perp} = C_T'(S_{10} - S_{11}\beta_0^*\alpha_0')\alpha_{0\perp} \Rightarrow \int_0^1 G(dW)'\alpha_{0\perp}$$

and

$$T^{-1}C_T' S_{11} C_T \Rightarrow \int_0^1 GG' du.$$

Noting that $S(\lambda) = S(\eta/T) = \eta T^{-1}S_{11} - S_{10}S_{00}^{-1}S_{01}$ and using the transformations (7.48) and (7.49), the roots of the equation

$$| \, C_T' S(\eta/T) C_T - C_T' S(\eta/T)\beta_0^*(\beta_0'^* S(\eta/T)\beta_0^*)^{-1}\beta_0'^* S(\eta/T) C_T \, |= 0$$

converge to those of the following equation

$$| \, \eta \int_0^1 GG' du - \int_0^1 G(dW)'\alpha_{0\perp}(Var(\alpha_{0\perp}' W))^{-1}\alpha_{0\perp}'\{\int_0^1 G(dW)'\}' \, |= 0. \qquad (7.50)$$

Let us define the following invertible matrix

$$J = \begin{pmatrix} (\bar\beta_{0\perp}' C\Sigma_\epsilon C'\bar\beta_{0\perp})^{-\frac{1}{2}} & 0 \\ 0 & 1 \end{pmatrix}.$$

Noting that

$$\begin{aligned} &(\alpha_{0\perp}' W)'(Var(\alpha_{0\perp}' W))^{-1}\alpha_{0\perp}' W = \\ &((\bar\beta_{0\perp}' C\Sigma_\epsilon C'\bar\beta_{0\perp})^{-\frac{1}{2}}\bar\beta_{0\perp}'\beta_{0\perp}(\alpha_{0\perp}'\Gamma_0\beta_{0\perp})^{-1}\alpha_{0\perp}' W)' \\ &(Var((\bar\beta_{0\perp}' C\Sigma_\epsilon C'\bar\beta_{0\perp})^{-\frac{1}{2}}\bar\beta_{0\perp}'\beta_{0\perp}(\alpha_{0\perp}'\Gamma_0\beta_{0\perp})^{-1}\alpha_{0\perp}' W))^{-1} \\ &(\bar\beta_{0\perp}' C\Sigma_\epsilon C'\bar\beta_{0\perp})^{-\frac{1}{2}}\bar\beta_{0\perp}'\beta_{0\perp}(\alpha_{0\perp}'\Gamma_0\beta_{0\perp})^{-1}\alpha_{0\perp}' W, \end{aligned}$$

and multiplying by J and J' the equation (7.50), the roots of (7.50) are the same of the following relation

$$| \, \eta \int_0^1 FF' du - \int_0^1 F(dB)'\{\int_0^1 F(dB)'\}' \, |= 0, \qquad (7.51)$$

where $B = (\bar\beta_{0\perp}' C\Sigma_\epsilon C'\bar\beta_{0\perp})^{-\frac{1}{2}}\bar\beta_{0\perp}' CW$ is such that $Var(B) = I_{d-r_0}$, and $F = (F_1, F_2)$ where $F_1 = B$ and $F_2 = u - \frac{1}{2}$. The equation (7.51) is equivalent to

$$| \, \eta I_{d-r_0+1} - \int_0^1 F(dB)'\{\int_0^1 F(dB)'\}'[\int_0^1 FF' du]^{-1} \, |= 0, \qquad (7.52)$$

so that denoting by η_i the eigenvalues of (7.52) we write

$$\sum_{r_0+1}^{d} \eta_i = tr\{\{\int_0^1 F(dB)'\}'[\int_0^1 FF' du]^{-1}\int_0^1 F(dB)'\}. \qquad (7.53)$$

Noting that as indicated above the roots of (7.45) are continuous functions of the matrices S_{11}, S_{10}, S_{00}, and S_{01}, we have

$$T\sum_{i=r_0+1}^{d} \hat\lambda_i \Rightarrow \sum_{r_0+1}^{d} \eta_i. \qquad (7.54)$$

Now writing the expression of the LR test statistic and since the the roots $\hat{\lambda}_{r_0+1},\ldots,\hat{\lambda}_d$ of (7.41) tends to zero at the rate T^{-1}, we find

$$
\begin{aligned}
-2\log Q_{r_0} = -T\sum_{i=r_0+1}^{d}\log(1-\hat{\lambda}_i) &= T[\sum_{i=r_0+1}^{d}\hat{\lambda}_i + o_p(T^{-1})] \\
&= T\sum_{i=r_0+1}^{d}\hat{\lambda}_i + o_p(1).
\end{aligned}
$$

Then using (7.54) and (7.53) the result follow. □

In order to prove Proposition 4.1 we have to state some additional asymptotic results. First note that in (7.17) multiplying by C'_T is equivalent (asymptotically) to multiplying by the transpose of

$$
\tilde{C}_T = \begin{pmatrix} \bar{\beta}_{0\perp} & 0 \\ 0 & T^{-\frac{1}{2}} \end{pmatrix}
$$

and suppose that the parameters in the deterministic part of (2.3) are equal to zero. To see this note that in this case the expression (2.2) becomes

$$
X_t = C\sum_{i=1}^{t}\epsilon_i + Y_t + A
$$

where Y_t is a stationary process, so that we have

$$
T^{-\frac{1}{2}}(\bar{\beta}'_{0\perp},0)Z_{1[Tu]} = T^{-\frac{1}{2}}\bar{\beta}'_{0\perp}C\sum_{i=1}^{[Tu]}\epsilon_i + T^{-\frac{1}{2}}\bar{\beta}'_{0\perp}Y_{[Tu]} + T^{-\frac{1}{2}}\bar{\beta}'_{0\perp}A. \tag{7.55}
$$

Therefore starting with (7.55) it is easy to see that one can retrieve the results of Lemma 7.5 and 7.6 replacing C_T by the new normalization matrix $\tilde{C}_T$. Then in the sequel we can assume without loss of generality that the parameters ν_0 and τ_0 are equal to zero. Now consider the following normalization of $\hat{\beta}^*$

$$
\tilde{\beta}^* = (\tilde{\beta}',\tilde{\tau})' = ((\hat{\beta}(\bar{\beta}'_{0c}\hat{\beta})^{-1})',((\hat{\beta}'\bar{\beta}_{0c})^{-1}\hat{\tau}))',
$$

where $\bar{\beta}_{0c} = \beta_{0c}(\beta'_{0c}\beta_{0c})^{-1}$ and define $\tilde{\alpha} = \hat{\alpha}\hat{\beta}'\bar{\beta}_{0c}$. Recall that $\hat{\tau}$ and $\tilde{\tau}$ are vectors of dimension r_0. For the rest of the paper we will use this normalization for theoretical derivations only since the matrix of unknown parameters β_{0c} appears in the expression of $\tilde{\beta}^*$. Note also that we take $\bar{\beta}_{0c}$ as a normalization matrix. Then in this case β^*_{0c} is the normalized matrix. With this notation and since we assumed $\tau_{0c}=0$, we have

$$
\tilde{\beta}^* = \beta^*_{0c} + \tilde{C}_T U_T \tilde{\beta}^* \tag{7.56}
$$

where

$$
U_T = \begin{pmatrix} \beta'_{0\perp} & 0 \\ 0 & T^{\frac{1}{2}} \end{pmatrix}.
$$

Note that (7.56) is obtained by projecting $\tilde{\beta}^*$ in the directions of β^*_{0c}, $\beta^*_{0\perp} = (\beta'_{0\perp},0)'$ and $\gamma = (0,1)'$, where γ is a vector of dimension $d+1$. Then it is seen from the d

first rows of (7.56) that with this choice of normalization $\tilde{\beta} - \beta_{0c}$ is included in the space spanned by $\beta_{0\perp}$. In the following Lemma we will state some asymptotic results we need using this normalization.

Lemma 7.8. *Under* **A1**, **A2** *and* **A3**, *we have*

$$\tilde{\alpha} \xrightarrow{P} \alpha_{0c}, \quad \hat{\Sigma}_\epsilon \xrightarrow{P} \Sigma_\epsilon, \quad \tilde{\beta} - \beta_{0c} = o_p(T^{-\frac{1}{2}}) \quad \text{and} \quad \tilde{\tau} - \tau_{0c} = o_p(T^{-1}).$$

Moreover the estimators $\tilde{\beta}$ and $\tilde{\tau}$ are such that

$$\begin{pmatrix} T\beta_{0\perp}'(\tilde{\beta} - \beta_{0c}) \\ T^{\frac{3}{2}}\tilde{\tau} \end{pmatrix} \Rightarrow [\int_0^1 GG'du]^{-1} \int_0^1 G(dV_\alpha)' \tag{7.57}$$

where

$$V_\alpha = (\alpha_{0c}'\Sigma_\epsilon^{-1}\alpha_{0c})^{-1}\alpha_{0c}'\Sigma_\epsilon^{-1}W$$

is independent of G.

Proof of Lemma 7.8. In a first time we will prove that $\tilde{\beta} - \beta_{0c} = o_p(T^{-\frac{1}{2}})$ and $\tilde{\tau} - \tau_{0c} = o_p(T^{-1})$. Let us define the matrix

$$B_T = \begin{pmatrix} \beta_{0c} & T^{-\frac{1}{2}}\bar{\beta}_{0\perp} & 0 \\ 0 & 0 & T^{-1} \end{pmatrix}.$$

Multiplying (7.41) by B_T' and B_T, we obtain

$$\mid B_T'(\lambda S_{11} - S_{10}S_{00}^{-1}S_{01})B_T \mid = 0. \tag{7.58}$$

Similarly to the proof of Proposition 3.1 and since we assumed that the deterministic terms are equal to zero, we have

$$\mid B_T'(S_{11} - S_{10}S_{00}^{-1}S_{01})B_T \mid \Rightarrow \begin{vmatrix} \lambda\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} & 0 \\ 0 & \lambda\int_0^1 GG'du \end{vmatrix}.$$

The eigenvectors g_i corresponding to the r_0 positive eigenvalues of the equation

$$\begin{vmatrix} \lambda\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} & 0 \\ 0 & \lambda\int_0^1 GG'du \end{vmatrix} = 0,$$

verify the equation

$$\begin{pmatrix} \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} & 0 \\ 0 & 0 \end{pmatrix} g_i = \begin{pmatrix} \lambda\Sigma_{\beta\beta} & 0 \\ 0 & \lambda\int_0^1 GG'du \end{pmatrix} g_i.$$

In addition the eigenvalues $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_r$ of (7.41) converge to those of the equation $\mid \lambda\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta} \mid = 0$, then it can be seen that the space spanned by the r_0 eigenvectors corresponding to the eigenvalues $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_r$ converges to the space spanned by the r_0 first unit vectors (the $d - r_0 + 1$ last coordinates of these eigenvectors converging to zero).

Thus since the eigenvectors of (7.58) are obtained by multiplying by B_T^{-1} the eigenvectors of (7.41) on the left, we write

$$B_T^{-1}\tilde{\beta}^* = \begin{pmatrix} I_r \\ T^{\frac{1}{2}}\beta'_{0\perp}\tilde{\beta} \\ T\tilde{\tau} \end{pmatrix} = \begin{pmatrix} I_r \\ o_p(1) \\ o_p(1) \end{pmatrix},$$

where B_T^{-1} is given by the following equation

$$B_T^{-1} = \begin{pmatrix} \bar{\beta}'_{0c} & 0 \\ T^{\frac{1}{2}}\beta'_{0\perp} & 0 \\ 0 & T \end{pmatrix} = \begin{pmatrix} \bar{\beta}'^*_{0c} \\ T^{\frac{1}{2}}U_T \end{pmatrix}.$$

Thus we can conclude that $\tilde{\tau} = o_p(T^{-1})$. In addition since $\tilde{\beta} - \beta_{0c}$ is included in the space of $\beta_{0\perp}$, we have $(\tilde{\beta} - \beta_{0c}) = o_p(T^{-\frac{1}{2}})$.

In this part of the proof we will show the consistency of $\tilde{\alpha}$ and $\hat{\Sigma}_\epsilon$. From Lemma (7.7) we have

$$\alpha_{0c} = \Sigma_{0\beta}\Sigma_{\beta\beta}^{-1} \quad \text{and} \quad \Sigma_\epsilon = \Sigma_{00} - \alpha_{0c}\Sigma_{\beta\beta}\alpha'_{0c}.$$

Since $U_T\tilde{\beta}^* = o_p(T^{-\frac{1}{2}})$, and using the relations (7.20) and (7.22) we have

$$\tilde{\beta}'^* S_{11}\tilde{\beta}^* = (\beta^*_{0c} + \tilde{C}_T U_T\tilde{\beta}^*)' S_{11}(\beta^*_{0c} + \tilde{C}_T U_T\tilde{\beta}^*) = \beta'^*_{0c} S_{11}\beta^*_{0c} + o_p(1). \tag{7.59}$$

Then from Lemma (7.7) we obtain

$$\tilde{\beta}'^* S_{11}\tilde{\beta}^* \xrightarrow{P} \Sigma_{\beta\beta}.$$

Similarly we have

$$\tilde{\beta}'^* S_{10} = \beta'^*_{0c} S_{10} + o_p(T^{-\frac{1}{2}}) \xrightarrow{P} \Sigma_{\beta 0}. \tag{7.60}$$

Finally writing the expressions of $\tilde{\alpha}$ and $\hat{\Sigma}$ we find

$$\tilde{\alpha} = S_{01}\tilde{\beta}^*(\tilde{\beta}'^* S_{11}\tilde{\beta}^*)^{-1} \xrightarrow{P} \alpha_{0c},$$

$$\hat{\Sigma}_\epsilon = S_{00} - S_{01}\tilde{\beta}^*(\tilde{\beta}'^* S_{11}\tilde{\beta}^*)^{-1}\tilde{\beta}'^* S_{10} \xrightarrow{P} \Sigma_\epsilon.$$

In order to prove the last statement of our Lemma, let us write the derivatives of the concentrated likelihood function (2.8) with respect to β^* in the direction h

$$\begin{aligned} D_{\beta^*}\log L(\alpha, \beta^*, \Sigma_\epsilon) &= \lim_{s\to 0}\frac{\log L(\alpha, \beta^* + sh, \Sigma_\epsilon) - \log L(\alpha, \beta^*, \Sigma_\epsilon)}{s} \\ &= Ttr\{\alpha'\Sigma_\epsilon^{-1}(S_{01} - \alpha\beta'^* S_{11})h\}. \end{aligned}$$

Noting that the matrices $\tilde{\alpha}$ and $\tilde{\beta}^*$ verifies the likelihood equation, this derivative is equal to zero at the point $(\tilde{\alpha}, \tilde{\beta}^*, \hat{\Sigma}_\epsilon)$ in all directions. Then we have

$$\tilde{\alpha}'\hat{\Sigma}_\epsilon^{-1}(S_{01} - \tilde{\alpha}\tilde{\beta}'^* S_{11}) = 0. \tag{7.61}$$

Recall that we have defined $N_\epsilon = T^{-1}\sum_{t=1}^{T} R_{1t}\epsilon_t'$. Inserting $S_{01} = \alpha_{0c}\beta_{0c}'^{*}S_{11} + N_\epsilon'$ in (7.61) we get

$$\begin{aligned}\tilde{\alpha}'\hat{\Sigma}_\epsilon^{-1}(S_{01} - \tilde{\alpha}\tilde{\beta}'^{*}S_{11}) &= \tilde{\alpha}'\hat{\Sigma}_\epsilon^{-1}(N_\epsilon' + \alpha_{0c}\beta_{0c}'^{*}S_{11} - \tilde{\alpha}\tilde{\beta}'^{*}S_{11})\\ &= \tilde{\alpha}'\hat{\Sigma}_\epsilon^{-1}(N_\epsilon' - \tilde{\alpha}(\tilde{\beta}^{*} - \beta_{0c}^{*})'S_{11} - (\tilde{\alpha} - \alpha_{0c})\beta_{0c}'^{*}S_{11}) = 0.\end{aligned}$$

Now multiplying by $\tilde{C}_T$ on the right and inserting $\tilde{\beta}'^{*} - \beta_{0c}^{*} = C_T U_T \tilde{\beta}^{*}$ we have

$$\tilde{\alpha}'\Sigma_\epsilon^{-1}(N_\epsilon'\tilde{C}_T - \tilde{\alpha}T\tilde{\beta}'^{*}U_T'\{T^{-1}\tilde{C}_T'S_{11}\tilde{C}_T\} - (\tilde{\alpha} - \alpha_{0c})\beta_{0c}'^{*}S_{11}\tilde{C}_T) = 0.$$

From the consistency of $\tilde{\alpha}$ and using (7.22) the last term tends to zero, so that we obtain

$$TU_T\tilde{\beta}^{*} = (T^{-1}\tilde{C}_T'S_{11}\tilde{C}_T)^{-1}\tilde{C}_T'N_\epsilon\Sigma_\epsilon^{-1}\alpha_{0c}(\alpha_{0c}'\Sigma_\epsilon^{-1}\alpha_{0c})^{-1} + o_p(1).$$

Finally using (7.20) and noting that from (7.31) and (7.32) we have $C_T'N_\epsilon \Rightarrow \int_0^1 G(dW)'$, we can deduce that

$$TU_T\tilde{\beta}^{*} \Rightarrow [\int_0^1 GG'du]^{-1}\int_0^1 G(dV_\alpha)'.$$

This complete the proof of Lemma (7.8). $\square$

Proof of Proposition 4.1. In a first time we will prove statement (4.1). From (7.57) we have

$$T\beta_{0\perp}'(\tilde{\beta} - \beta_{0c}) \Rightarrow [\int_0^1 G_{1.2}G_{1.2}'du]^{-1}\int_0^1 G_{1.2}(dV_\alpha)'.$$

From the d first rows of (7.56) we write

$$\tilde{\beta} - \beta_{0c} = \bar{\beta}_{0\perp}\beta_{0\perp}'(\tilde{\beta} - \beta_{0c}).$$

Then using the expansion

$$(\hat{\beta}_c - \beta_{0c}) = (I_d - \beta_{0c}c')(\tilde{\beta} - \beta_{0c}) + O_p(\| (\tilde{\beta} - \beta_{0c}) \|^2) \tag{7.62}$$

and noting that since $\tilde{\beta} - \beta_{0c}$ is included in the space of $\beta_{0\perp}$ we have $\| (\tilde{\beta} - \beta_{0c}) \|^2 = O_p(T^{-2})$ the result follow. Similarly writing $\tilde{\tau} = (\hat{\beta}'\bar{\beta}_{0c})^{-1}(\hat{\beta}'c)\hat{\tau}_c$, we can find that $\hat{\tau}_c = \tau_{0c} + O_p(T^{-\frac{3}{2}})$. Now let W_1 and W_2 two independent Brownian motions. The form (4.2) can be found noting that given W_1, $\int_0^1 W_1(dW_2)'$ is gaussian with mean zero and variance matrix

$$\int_0^1 W_1W_1' \otimes Var(W_2). \qquad \square$$

Recall that $\hat{\alpha}_c(\beta_{0c}^{*}) = S_{01}\beta_{0c}^{*}(\beta_{0c}'^{*}S_{11}\beta_{0c}^{*})^{-1}$ and $\hat{\alpha}_c = S_{01}\hat{\beta}_{0c}^{*}(\hat{\beta}_{0c}'^{*}S_{11}\hat{\beta}_{0c}^{*})^{-1}$. To prove Proposition 4.2 we need to state the following Lemma.

Lemma 7.9. *Under* **A1**, **A2** *and* **A3**, *we have*

$$\hat{\alpha}_c = \hat{\alpha}_c(\beta_{0c}^{*}) + o_p(T^{-\frac{1}{2}}).$$

Proof of Lemma 7.9. First note that we have

$$\begin{aligned}
\tilde{\alpha} = \hat{\alpha}\hat{\beta}'_{0c}\bar{\beta}_{0c} &= S_{01}\hat{\beta}^*_{0c}(\hat{\beta}'^*_{0c}S_{11}\hat{\beta}^*_{0c})^{-1}\hat{\beta}'_{0c}\bar{\beta}_{0c} \\
&= S_{01}\hat{\beta}^*_{0c}(\bar{\beta}'_{0c}\hat{\beta}_{0c})^{-1}(\bar{\beta}'_{0c}\hat{\beta}_{0c})(\hat{\beta}'^*_{0c}S_{11}\hat{\beta}^*_{0c})^{-1}(\bar{\beta}'_{0c}\hat{\beta}_{0c})' \\
&= S_{01}\tilde{\beta}^*_{0c}(\tilde{\beta}'^*_{0c}S_{11}\tilde{\beta}^*_{0c})^{-1}.
\end{aligned}$$

From (7.59) and (7.60) we obtain

$$\tilde{\alpha} = \hat{\alpha}_c(\beta^*_{0c}) + o_p(T^{-\frac{1}{2}}). \tag{7.63}$$

Recall that $\hat{\alpha}_c\hat{\beta}'_c = \tilde{\alpha}\tilde{\beta}'$. Noting that $\beta'_{0c}c = \hat{\beta}'_c c = I_r$, we write

$$\begin{aligned}
\hat{\alpha}_c &= \tilde{\alpha}\tilde{\beta}'c \\
&= \tilde{\alpha}(\tilde{\beta} - \beta'_{0c})c + \tilde{\alpha}\beta'_{0c}c \\
&= \tilde{\alpha}(\tilde{\beta} - \beta'_{0c})c + \tilde{\alpha}.
\end{aligned}$$

In view of the consistency of $\tilde{\alpha}$ and since $\tilde{\beta} = \beta_{0c} + O_p(T^{-1})$, we have

$$\hat{\alpha}_c = \tilde{\alpha} + O_p(T^{-1}),$$

and then the result follow from (7.63). $\Box$

Proof of Proposition 4.2. Multiplying (2.9) by $R'_{1t}\beta^*_{0c}$ on the right we find

$$\begin{aligned}
\alpha_{0c} &= T^{-1}\sum_{t=1}^{T}(R_{0t} - \epsilon_t)R'_{1t}\beta^*_{0c}(\beta'^*_{0c}S_{11}\beta^*_{0c})^{-1} \\
&= T^{-1}\sum_{t=1}^{T}(R_{0t} - \epsilon_t)R'_{1t}\beta^*_{0c}(\beta'^*_{0c}S_{11}\beta^*_{0c})^{-1}.
\end{aligned}$$

Then from Lemma 7.9 and using (2.10) and (7.33) we have

$$\begin{aligned}
T^{\frac{1}{2}}vec(\hat{\alpha}_c - \alpha_{0c}) &= T^{\frac{1}{2}}vec(\hat{\alpha}_c - \hat{\alpha}_c(\beta^*_{0c})) + T^{\frac{1}{2}}vec(\hat{\alpha}_c(\beta^*_{0c}) - \alpha_{0c}) \\
&= T^{-\frac{1}{2}}\sum_{t=1}^{T}vec(\epsilon_t R'_{1t}\beta^*_{0c}(\beta'^*_{0c}S_{11}\beta^*_{0c})^{-1}) + o_p(1) \\
&= T^{-\frac{1}{2}}\sum_{t=1}^{T}\{(\beta'^*_{0c}S_{11}\beta^*_{0c})^{-1}\beta'^*_{0c}R_{1t} \otimes I_d\}\epsilon_t + o_p(1) \\
&= T^{-\frac{1}{2}}\sum_{t=1}^{T}\Sigma_c^{-1}\beta'^*_{0c}\tilde{R}_{1t} \otimes \epsilon_t + o_p(1) \\
&= (\Sigma_c^{-1} \otimes I_d)T^{-\frac{1}{2}}\sum_{t=1}^{T}\beta'^*_{0c}\tilde{R}_{1t} \otimes \epsilon_t + o_p(1) \\
&= (\Sigma_c^{-1} \otimes I_d)T^{-\frac{1}{2}}\sum_{t=1}^{T}v_t + o_p(1)
\end{aligned}$$

where

$$v_t = vec(\epsilon_t \tilde{R}'_{1t} \beta^*_{0c}).$$

Recall that we have defined

$$\tilde{R}_{1t} = Z_{1t} - \tilde{M}_{12}\tilde{M}_{22}^{-1} Z_{2t}.$$

Then using (7.24) and (7.25) it is easy to see that $\beta'^*_{0c}\tilde{R}_{1t}$ can be written as follows

$$\beta'^*_{0c}\tilde{R}_{1t} = m + \sum_{i=0}^{\infty} \tilde{\psi}_i \epsilon_{t-i-1}, \tag{7.64}$$

where m is a vector of constants and the terms of the series $\{\tilde{\psi}_i\}_{i\in\mathbb{N}}$ decay exponentially fast. Then despite the fact that there is a constant in the expression (7.64), we can show following the same lines of the proof of Lemma 7.4 that $T^{-\frac{1}{2}}\sum_{t=1}^{T} v_t$ is normally distributed. The form of the matrix Σ_α is obtained from the following computations

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{cov}(v_t, v_s) = \lim_{T\to\infty} \frac{1}{T}\sum_{|h|<T} (T- \mid h \mid)\mathrm{cov}(v_t, v_{t-h})$$

$$= \sum_{h=-\infty}^{\infty} \mathrm{cov}\,(v_t, v_{t-h}) = \sum_{h=-\infty}^{\infty} E\left\{\beta'^*_{0c}\tilde{R}_{1t}\otimes\epsilon_t\right\}\left\{\beta'^*_{0c}\tilde{R}_{1t-h}\otimes\epsilon_{t-h}\right\}'.$$

Finally we obtain

$$\Sigma_\alpha = \sum_{h=-\infty}^{\infty} E\left\{\Sigma_c^{-1}\beta'^*_{0c}\tilde{R}_{1t}\tilde{R}'_{1t-h}\beta^*_{0c}\Sigma_c^{-1}\otimes\epsilon_t\epsilon'_{t-h}\right\},$$

using the well known identity $(A\otimes B)(C\otimes D) = AC\otimes BD$. □

References

Ahn, S.K. and Reinsel, G.C. (1990) Estimation for partially nonstationary multivariate autoregressive models. *Journal of American Statistical Association* 100, 813-823.

Anderson, T.W. (1951) Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22, 327-351.

Andrews, B., Davis, R.A., and Breidt, F.J., (2006) Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis* 97, 1638-1659.

Breidt, F.J., and Davis, R.A., Trindade, A. (2001) Least absolute deviation estimation for all-pass time series models. *The annals of statistics* 29, 919-946.

Clements, M.P., and Hendry, D.F. (1996) Intercept corrections and structural change. *Journal of Applied Econometrics* 11, 475-494.

Davydov, Y. A. (1968) Convergence of distributions generated by stationary stochastic process. *Theory of Probability and Applications* 13, 691-696.

ENGLE, R.F. AND GRANGER, C.W.J. (1987) Co-integration and error correction: representation, estimation, and testing. *Econometrica* 55, 251-276.

FRANCQ, C., AND RAÏSSI, H. (2005) Multivariate portmanteau test for autoregressive models with uncorrelated but nonindependent errors. Working document. Université Lille 3.

FRANCQ, C., ROY, R. AND ZAKOÏAN, J-M. (2005) Diagnostic checking in ARMA models with uncorrelated errors. *Journal of American Statistical Association* 100, 532–544.

GRANGER, C.W.J. (1981) Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* 16, 121-130.

HERRNDORF, N. (1984) A functional central limit theorem for weakly dependent sequences of random variables. *Annals of Probability* 12, 141-153.

JEANTHEAU, T. (1998) Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* 14, 70-86.

JOHANSEN, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231-254.

JOHANSEN, S. (1991) Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica* 59, 1551-1580.

JOHANSEN, S. (1995) *Likelihood-based inference in cointegrated vector autoregressive models.* New York: Oxford University Press.

JOHANSEN, S. AND JUSELIUS, K. (1990) Maximum likelihood estimation and inference on cointegration: with applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 52, 169-210.

LÜTKEPOHL, H. AND SAIKKONEN, P. (1999) A lag augmentation test for the cointegrating rank of a VAR process. *Economics Letters* 63, 23-27.

PERRON, P. AND CAMPBELL, J.Y. (1993) A note on Johansen's cointegration procedure when trend are present. *Empirical Economics* 18, 777-789.

PHILLIPS, P.C.B. (1988) Weak convergence to the stochastic integral $\int_0^1 BdB'$. *Journal of Multivariate Analysis* 24, 252-264.

PHILLIPS,P.C.B., AND DURLAUF, S.N. (1986) Multiple time series regression with integrated processes. *The Review of Economic Studies* 53, 473-495.

PHILLIPS,P.C.B., AND SOLO, V. (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 970-1001.

RAHBEK, A., HANSEN, E., DENNIS, J.G. (2002) ARCH innovations and their impact on cointegration rank testing. Working document. University of Copenhagen.

ROMANO, J. P. AND THOMBS, L. A. (1996) Inference for autocorrelations under weak assumptions. *Journal of American Statistical Association* 91, 590-600.

SEO, B. (2007) Asymptotic distribution of the cointegrating vector estimator in error correction models with conditional heteroskedasticity. *Journal of Econometrics* 137, 68-111.

TRENKLER, C. (2003) The polish exchange rate system: A unit root and cointegration analysis. *Empirical Economics* 28, 839-860.

Tables and Figures

TABLE 1: Empirical size (in %) of the LR test for VECM (5.4) in the strong and weak cases with $T = 100$.

	case 1	case 2	case 3
MD (5.1)	**8.2**	**6.6**	**1.6**
MD (5.2)	**6.5**	6.3	5.3
WWN (5.3)	**3.3**	4.4	**0.6**
SWN	5.2	5.2	**2.6**

Parameters: $\pi_2 = 0.9$ $e = -1$ $\theta = -1.5$. Case 1: $\pi_1 = -0.1$ and $e\pi_2 + \pi_1 = -1$. Case 2: $\pi_1 = 0.8$ and $e\pi_2 + \pi_1 = -0.1$. Case 3: $\pi_1 = -0.8$ and $e\pi_2 + \pi_1 = -1.7$.

TABLE 2: As Table 1, but for $T = 400$.

	case 1	case 2	case 3
MD (5.1)	5.8	4.2	5.8
MD (5.2)	5.2	5.5	**6.4**
WWN (5.3)	4.0	5.0	**2.9**
SWN	5.0	4.6	5.0

TABLE 3: The relative rejection frequencies (in %) of the LR test for VECM (5.4) with heteroscedastic errors (5.5).

f	0	0.005	0.01	0.015	0.02
$T = 100$	5.2	3.1	2.1	1.1	0.8
$T = 400$	5.0	2.3	2.0	0.8	0.2

TABLE 4: Empirical power (in %) of the LR test for the $AR(1)$ model (5.6) in the strong and weak case with $T = 100$ and $e\pi_2 + \pi_1 = -0.85$.

ϖ	0	-0.05	-0.1	-0.15	-0.2	-0.25	-0.3	-0.35
MD (5.1)	7.4	20.3	35.9	55.5	74.0	86.5	93.8	98.1
MD (5.2)	6.3	18.8	36.9	57.6	78.3	90.8	96.6	99.0
WWN (5.3)	3.8	13.9	28.6	52.1	75.6	90.7	97.8	99.5
SWN	4.8	14.8	30.7	55.0	78.2	93.2	97.9	100.0

Case: $\pi_1 = -0.7 \quad \pi_2 = 0.15 \quad e = -1 \quad \theta = -1.5$.

TABLE 5: As Table 4, but for $T = 400$.

ϖ	0	-0.05	-0.1	-0.15	-0.2
MD (5.1)	6.3	72.2	99.7	100.0	100.0
MD (5.2)	6.2	74.4	100.0	100.0	100.0
WWN (5.3)	4.7	72.6	99.8	100.0	100.0
SWN	5.9	74.3	99.9	100.0	100.0

TABLE 6: As Table 4, but for $e\pi_2 + \pi_1 = -1.8 \approx -2$.

ϖ	0	-0.03	-0.04	-0.05	-0.06	-0.07	-0.09	-0.11
MD (5.1)	0.3	16.0	32.2	52.4	67.8	80.3	91.6	99.8
MD (5.2)	5.5	7.1	20.6	43.8	65.8	81.6	94.3	98.9
WWN (5.3)	0.1	18.3	33.6	52.6	69.0	80.3	94.1	98.1
SWN	1.3	13.1	30.4	51.2	67.8	82.2	95.4	98.8

Case: $\pi_1 = -0.9 \quad \pi_2 = 0.9 \quad e = -1 \quad \theta = -1.5$.

TABLE 7: As Table 6, but for $T = 400$.

ϖ	0	-0.03	-0.04	-0.05	-0.06	-0.07	-0.09	-0.11
MD (5.1)	4.9	0.5	20.2	72.1	96.3	99.9	100.0	100.0
MD (5.2)	6.1	0.0	2.8	62.4	97.3	99.9	100.0	100.0
WWN (5.3)	2.2	3.7	27.6	72.5	96.9	99.9	100.0	100.0
SWN	4.3	0.3	22.9	74.4	98.0	99.9	100.0	100.0

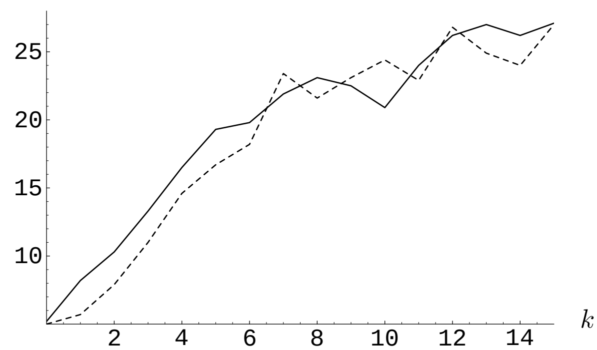

FIGURE 7.1: The relative rejection frequencies (in %) of the LR test for different values of k in the weak white noise (5.1) for $T = 100$ (full line) and $T = 400$ (dotted line). Case 1: $\pi_1 = -0.1 \quad \pi_2 = 0.9$ $e = -1 \quad \theta = -1.5$. Number of replications $n = 1000$.

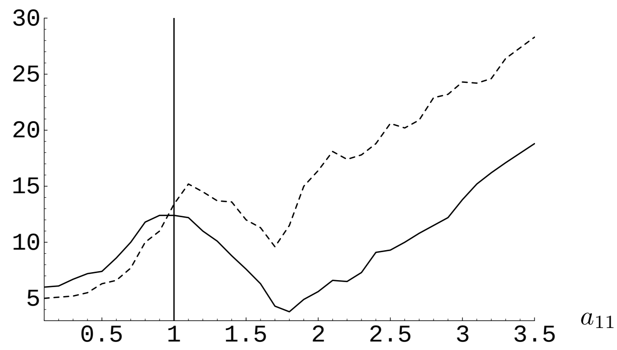

FIGURE 7.2: The same as in Figure 7.1 with a weak white noise which follow an ARCH model (5.2) with $a_{11} = a_{22}$ and $a_{21} = a_{12} = 0$.

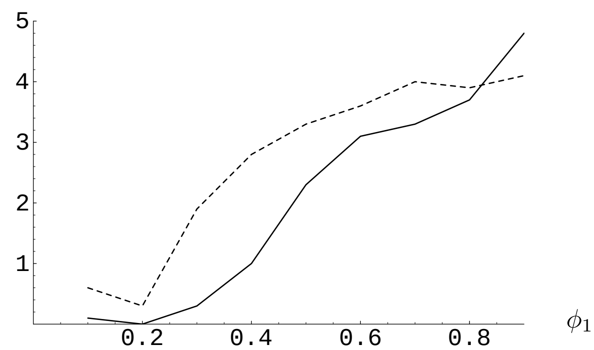

FIGURE 7.3: The same as in Figure 7.1 with a weak white noise which follow an all-pass model (5.3) with $\phi_1 = \phi_2$.

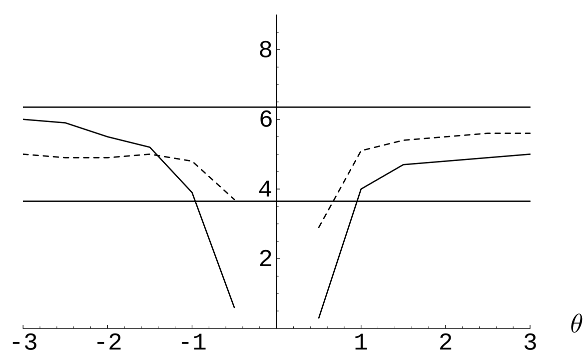

FIGURE 7.4: Effect of the trend parameter θ in the strong case: Relative rejection frequencies (in %) of the LR test for different values of θ in model (5.4) with iid errors for $T = 100$ (full line) and $T = 400$ (dotted line). Case: $\pi_1 = -0.1 \quad \pi_2 = 0.9 \quad e = -1$. Number of replications $n = 1000$.

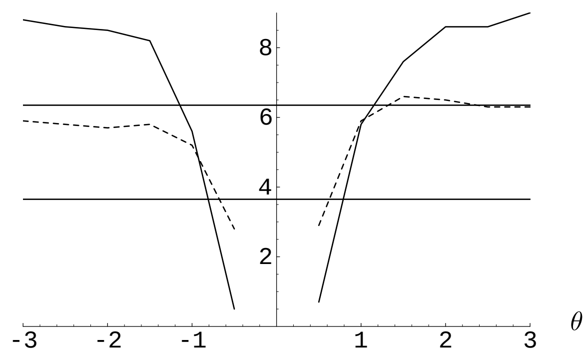

FIGURE 7.5: Effect of trend parameter θ, the weak white noise (5.1) case: The same as in Figure 7.4 but for an error process which follow (5.1) with $k = 1$.

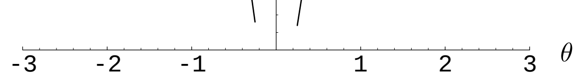

FIGURE 7.6: Effect of trend parameter θ, the ARCH case: The same as in Figure 7.4 but for an error process which follow (5.2) with $a_{12} = a_{21} = 0.1$, $a_{11} = 0.2$ and $a_{22} = 0.3$.

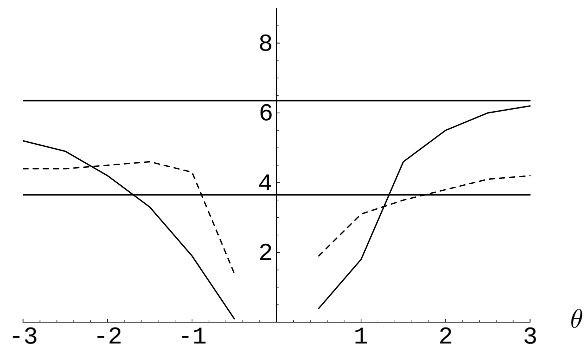

FIGURE 7.7: Effect of trend parameter θ, the all-pass case: The same as in Figure 7.4 but for an error process which follow (5.3) with $\phi_1 = \phi_2 = 0.7$ for $T = 100$ (full line) and $T = 800$ dotted line.

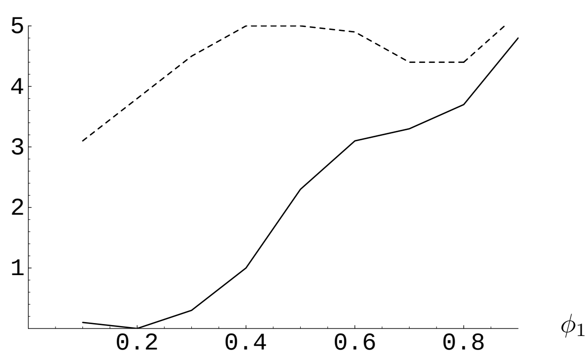

FIGURE 7.8: The relative rejection frequencies (in %) of the LR test for different values of $\phi_1 = \phi_2$ in weak white noise (5.3) for $T = 100$. Case1: $e\pi_2 + \pi_1 = -1$ (full line). Case 2: $e\pi_2 + \pi_1 = -0.1$ (dotted line). Number of replications $n = 1000$.

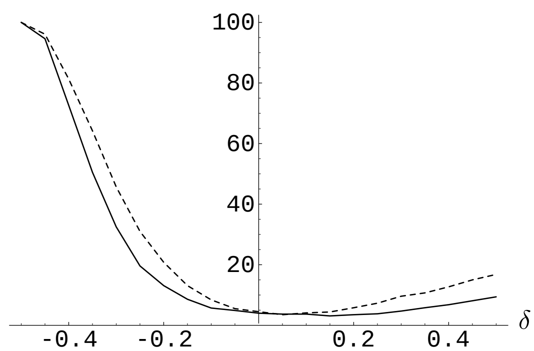

FIGURE 7.9: The relative rejection frequencies (in %) of the LR test with correlated errors for $T = 100$ (full line) and $T = 400$ (dotted line). Case: $\pi_1 = 0.9 \quad \pi_2 = -1 \quad e = 1 \quad \theta = -0.5$. Number of replications $n = 1000$.

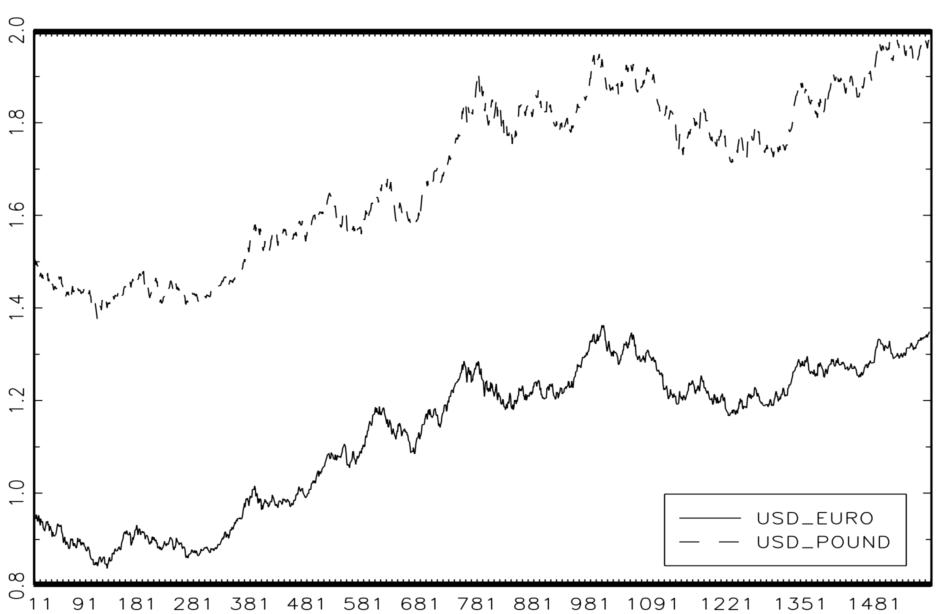

FIGURE 7.10: The daily exchange rates of U.S. Dollars to one British Pound and of U.S. Dollars to one Euro. Data source: The Research Division of the Federal Reserve Bank of St. Louis www.research.stlouisfed.org.

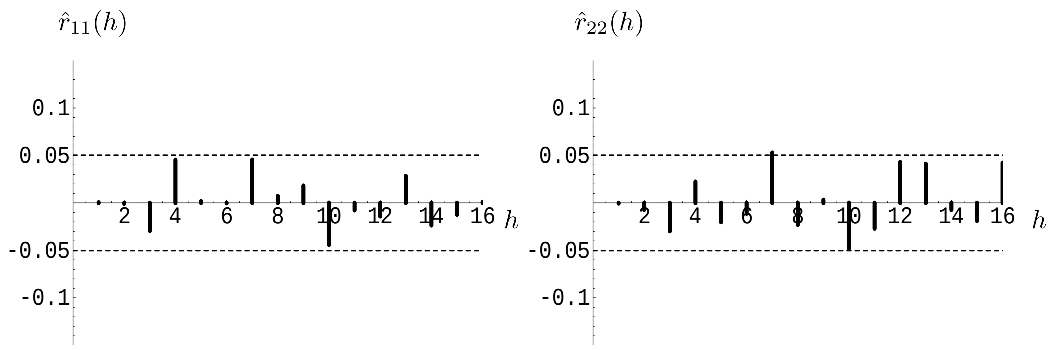

FIGURE 7.11: Autocorrelations of the residuals of the VECM with $r_0 = 1$ and $p = 2$ for the the daily exchange rates of U.S. Dollars to one British Pound and of U.S. Dollars to one Euro. The left graphic represent the autocorrelations $\hat{r}_{11}(h)$ of the residuals $\hat{\epsilon}_{1t}$ and the right the autocorrelations $\hat{r}_{22}(h)$ of the residuals $\hat{\epsilon}_{2t}$. The horizontal lines about zero represent the approximate 5% significance limits for the sample autocorrelations (that is $\pm 1.96/\sqrt{T}$ with $T = 1578$).

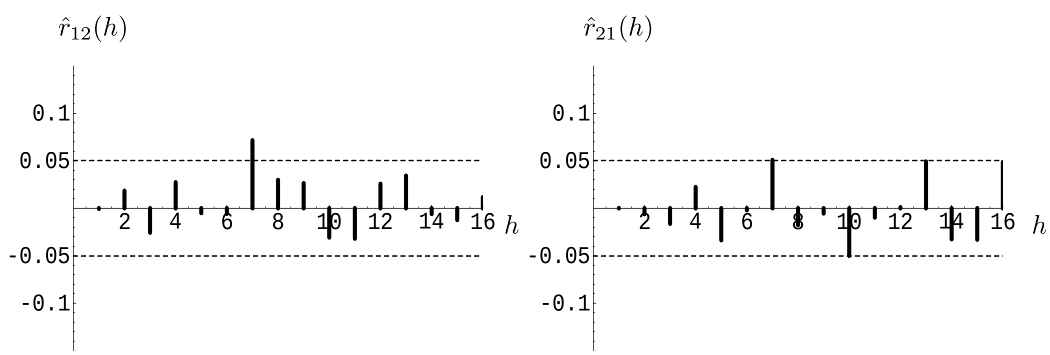

FIGURE 7.12: The same as for the Figure 7.11 but for the crosscorrelations of the $\hat{\epsilon}_{1t}$'s and the $\hat{\epsilon}_{2t}$'s with obvious notations.

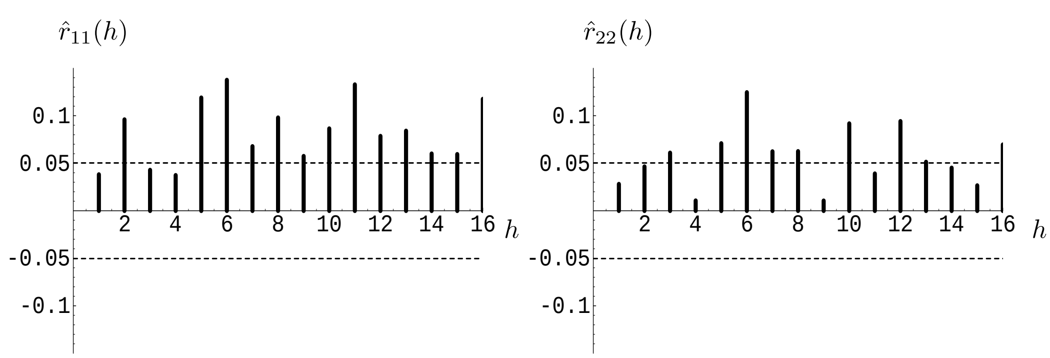

FIGURE 7.13: The same as for the Figure 7.11 but for squared residuals of the analyzed series.

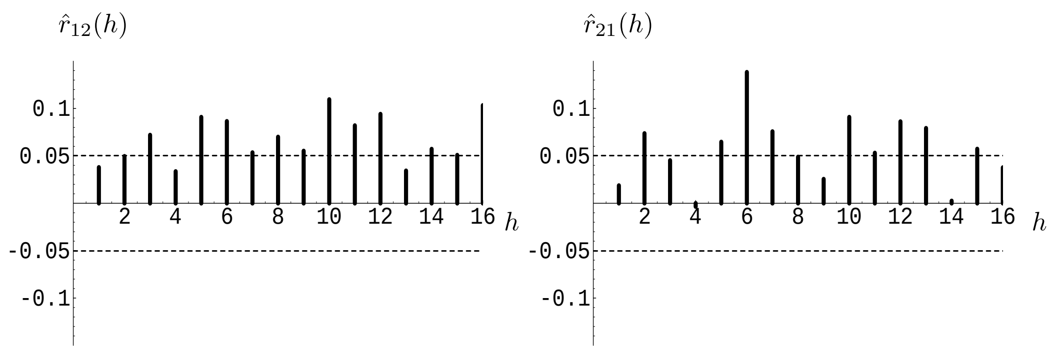

FIGURE 7.14: The same as for the Figure 7.12 but for squared residuals of the analyzed series.